Book of papers

June 8-10, DTU, Denmark
# Table of Contents

## Transport sustainability and environment

**Cars longevity: a biometric approach.** ZÉHIR KOLLI (INRETS / DEST, FRANCE)

**Impact of the penetration rate of ecodriving on fuel consumption and traffic congestion.** OLIVIER ORFILA (LCPC, FRANCE)

**Monitoring of the landslides on the Pan-European corridor X in the purpose of environmental protection.** ZORAN BERISAVLJEVIC (THE HIGHWAY INSTITUTE, SERBIA)

**Oil spill risk analysis of routeing heavy ship traffic in Norwegian waters.** JUNED AKHTAR (TØI, NORWAY)

**Bringing transparency to environmental impacts of transport- LIPASTO transport emission database.** HEIDI AUVINEN (VTT, FINLAND)

**Photocatalytic roads: from lab tests to real scale applications.** ELIA BOONEN (BRRC, BELGIUM)

**Driving in the future: What influences young males and females driving intentions?** SIGRUN BIRNA SIGURDARDOTTIR (DTU TRANSPORT, DENMARK)

## Transport Safety

**Very severely injured: In-depth investigation of road accident characteristics and medical consequences in Germany.** EIKE A. SCHMIDT (BAST, GERMANY)

**Measures for improving road safety on rural roads in Germany - (“AOSI”).** THOMAS JÄHRIG (BAST, GERMANY)

**Older drivers’ self-regulation in traffic.** ANNETTE MENG (DTU TRANSPORT, DENMARK)

**Quantifying the influence of social characteristics on accident and injuries risk: A comparative study between motorcyclists and car driver.** ALLAN LYCKEGAARD (DTU TRANSPORT, DENMARK), MORTEN N. OLESEN (ATKINS, DENMARK), TOVE HELS (DTU TRANSPORT, DENMARK)

**Quantifying driver distraction- the case study of Thessaloniki’s ring road.** KATERINA CHRYSOSTOMOU (HIT, GREECE)

**Road risk as a presidential priority: a public action for safety and a political instrument.** HUGUES CUNEGATTI (INRETS / DEST, FRANCE)

**Effect of thin water film on tire/road friction.** Y. BEAUTRU (LCPC, FRANCE), M. KANE (LCPC, FRANCE), V. CEREZO (LCPC, FRANCE), M. DO (LCPC, FRANCE)

**The development of Roadside Safety Criteria for Portuguese Roads.** CARLOS ROQUE (LNEC, PORTUGAL)

**Evaluation of the road safety – based incentives in Public Private Partnerships (PPPs): The case of Spain.** THAIS RANGEL (TRANSyT, SPAIN)
Effect of tyre tread depth on accident involvement during summer in Finland. Riikka Rajamäki (VTT, Finland)

Recognising transitions between rural road categories: the role of road layout and intersection type. Agnieszka Stelling (SWOV, The Netherlands)

Transport Economics, Policy and Behaviour

Economic costs of road traffic accidents in Germany. Thomas Kranz (BAST, Germany)

Geographical location of depopulation areas in the Czech Republic and its dependence on transport infrastructure. Emil Drápeľa (CDV, Czech Republic)

Elasticity of long distance travelling. Mette Aagaard Knudsen (DTU Transport, Denmark)

Improving the analysis of a toll ring scheme implementation by a Travel Demand Management model. Wang Yang (TRANSYT, Spain)

Competitive tendering in an entry regulated market – an accident waiting to happen? Jørgen Aarhaug (TØI, Norway)

Determinants of capacity utilization in road freight transportation. Megersa Abate (DTU Transport)

Evaluation of policy packages: Lessons learned from Bergen. Petter Christiansen (TØI, Norway)

Transport, Civil and Road Engineering

Weigh-in-motion on German motorways. Stefan Tetzner (BAST, Germany)

Integral consideration of the lightweight design for railway vehicle. Jens Koenig (DLR, Germany)

Early increases of noise from newly laid road surface. Jens Oddershede (Road Directorate/Danish Road Institute, Denmark)

A better approach to predict subsidence due to ground water leakage into tunnel. Vikas Thakur (The Norwegian Public Roads Administration, Norway)

Bridge rehabilitation. Dejan Cerovac (The Highway Institute, Serbia)

The relationship between road design and driving behavior: a simulator study. Liva Abele (DTU Transport, Denmark)

Evolution of pavement friction: Discerning the effects of load, velocity and aggregate types on polishing of pavements. Dan Zhao (LCPC, France), Malal Kane (LCPC, France), Minh- Tan Do (LCPC, France), François De Larrard (LCPC, France)

Sustainable maintenance of rural roads in Slovakia. Lubomir Pepucha (University of Zilina, Slovakia), Martin Pitonak (University of Zilina, Slovakia)

Relationship between road width and safety. Jiří Ambros (CDV, Czech Republic)
Intelligent Transport Systems & Traffic

Emergency vehicle prioritization using vehicle-to – infrastructure communication. LAURA BIEKER (DLR, GERMANY)

The impact of traffic management operations on motorway congestion. A level of service based evaluation method. JUDITH PRINCETON (INRETS, FRANCE)

External factors affecting motorway capacity. BEN MORRIS (TRL, UNITED KINGDOM)

Reliability of in-vehicle warning system for railway level crossing – a user- oriented analysis. RISTO ÖÖRNI (VTT, FINLAND)

Methodods for assessing the pedestrian level of service: International experience and adjustment to the Greek walking environment – The case of Thessaloniki. LEFTERIS SDOUKOPOULOS (CERTH/HIT, GREECE)

Optimization of key design elements of 2+1-route. MARCO IRZIK (BAST, GERMANY)

Evaluating network wide traffic management in the Amsterdam region. SUERD POLDERDIJK (RIJKSWATERSTAAT CENTRE FOR TRANSPORT AND NAVIGATION, THE NETHERLANDS)

Perception, change detection and utilization of roads using dynamic information: a theoretical review. ILSE M. HARMS (RIJKSWATERSTAAT CENTRE FOR TRANSPORT AND NAVIGATION, MINISTRY OF TRANSPORT, THE NETHERLANDS)
CAR LONGEVITY: A BIOMETRIC APPROACH

Zéhir KOLLI DEST /IFSTTAR
2 rue de La Butte Verte, 93166 Noisy le Grand, France
+33 1 45 94 55 96 zehir.kolli@ifsttar.fr

Abstract:
Vehicle longevity and survival are important to forecast total fleets and to estimate related greenhouse gas (GHG) emissions. Despite today’s abundant literature,¹ the major issue of the estimated survival rate of vehicles as a tool for vehicle and emissions forecasts is still an open question largely because of no data on scrapped cars. To secure an overview of car longevity, we applied a biometric approach by building Kaplan-Meier survival curves. Our database covers 265 vehicles² scrapped or sold for spare parts in France in August 2010. The information includes make, model, version, cylinder capacity, fiscal horsepower, fuel (diesel/gasoline), date of registration, date of scrapping and odometer reading.

Using the age in total months and final odometer readings of scrapped cars, we applied the Kaplan-Meier estimator to obtain survival rates by age and total mileage. The distinction between these two dimensions is useful to separate longevity by age (lifetime) from longevity by total mileage (design life). We found longer lifetime for gasoline cars but higher design life for diesel cars. The results give a median lifetime of 126 months (10y6m) years for the dataset, i.e. 149 months (12y5m) for gasoline cars and 118 months (9y10m) for diesel. Moreover, results give a median design life of 150,000 km (94,000 miles) for the dataset, i.e. 135,479 km (84,674 miles) for gasoline cars and 152,730 km (95,456 miles) for diesel. However, our study also shows that Kaplan-Meier survival curves tie in nicely with a two-parameter Weibull survival function.

Keywords: vehicle scrapping, biometric, vehicle longevity, Kaplan-Meier estimator, Weibull, lognormal, loglogistic, vehicle survival rates, vehicle retention rates

² V Data was compiled from a field survey plus various websites and follow-up calls for missing details.
1. INTRODUCTION

1.1. Background

Air pollution and GHG emissions from vehicles is no new issue (e.g. Southworth 1986; Zachariadis et al. 1995) and relevant critical issues are fleet size, usage (annual mileage), survival rates and lead time to deployment of new technologies (Zachariadis et al. 1995).

In this context, the evolution and internal dynamics of the passenger car market are of interest to numerous institutions and planners concerned with energy demand, resource consumption, public transport and traffic infrastructure planning, GHG and other emissions forecasting, not to mention the auto industry. To assess this growth and its dynamics, we looked at the following four large families of statistical models to determine car fleet size and composition:

- Aggregate models (Tanner 1958 and 1978) where forecasts of the total vehicle fleet were extrapolated from past and current trends of car ownership, combined with hypotheses concerning the evolution of the total fleet. This family of models includes Gompertz curves which obtain fleet size but not composition.
- Discrete choice models (Train 1978) to connect disaggregate analysis of durable ownership to the utility concept underpinning demand theory. These models are typically referred to as vehicle holding/transaction models.
- Demographic models (Madre 1989) based on longitudinal analysis of behavior. Longitudinal analysis is a demographic method used to interpret the effects of natality, marriage, mortality and other key transversal indicators by comparing them to the figures for successive age groups.
- Survival or scrappage models (Parks 1977; Berkovec 1985; Zachariadis et al. 1995 Greenspan and Cohen 1996) which involve a system-dynamic approach including car registration statistics that are readily available and estimations of car scrappage through survival data analysis. The model had to combine the replacement of old cars with new ones due to endogenous aging and "death" on the car market.

1.2. Previous work

A better understanding of fleet composition is important to design and implement policies and programs designed to stimulate new car sales and the retirement of gas-guzzlers, e.g. trade-in incentives and feebates. The average age of vehicles turned in under the 2009 French “Cash for Clunkers” trade-in incentive program was 15 years and 7 months.3 The use of smog checks, safety checks and other maintenance data for modeling the vehicle lifetimes ignores the disposal of vehicles due to accidents. Therefore, vehicle lifetimes and total fleet sizes are often overestimated.

Vehicle longevity suffers measurement in the following ways:

- Design life is based on vehicle scrappage due to mechanical breakdown or deterioration and smog-check data is one basis for this estimation.

3 Source: “Service d’Observation et des Statistiques” (SOeS 2010).
“Scrappage rate” is effective scrappage due to engineering failure or physical deterioration or other breakdown as well as road accidents.

“Imported and exported vehicles rates” is the annual proportion of imported and exported second-hand cars to new car registrations.

“Cyclical scrappage rates” refers to non-mechanical related vehicle scrappage, mostly due to trade-in incentives programs.

Researchers have approached these considerations from several standpoints with different methods. Much existing research for the US market covers longevity, survival or trade-in incentive programs and uses a logistic functional form to estimate survival rates per year or by vintage. Greene and Chen (1981) used a logistic equation separately for domestic and imported U.S. cars, trucks and light-duty vehicles from 1966 to 1977. Their results show better survival expectancy for light trucks than light vehicles, with medians of 14.5 and 9.9 years respectively.

Berkovec (1985) produced a scrappage model from log regression of the death rate on auto prices, prices squared and dummy variables for vintage and class. Libertiny (1993) applied a Weibull (1965) distribution to calculate attrition rates of passenger cars and found no significant difference between domestic and imported cars. Libertiny also found that while vehicle scrappage rates plunged between 1970 and 1980, they were relatively flat from 1980 to 1990. Greenspan and Cohen (1996) estimated car mortality to study the demand for new cars. They captured scrappage rates with a pooled regression. For consistency, they assumed zero scrappage for the first three years.

Gallez (2000) fitted longitudinal annual survival rates on French panel data by applying the assumption that they follow a lognormal survival function up to a maximum age of 28 years, after which he posits zero survival. Using data for France from 1968 to 1994, Gallez found median lifetimes of 10.82 and 11.23 years for diesel and gasoline cars respectively against corresponding mean lifetimes of 11.74 and 12.22 years for light vehicles in 1994.

Chen and Niemeier (2004) applied a Cox proportional hazard model of the Weibull form with two mass points. Odometer readings were introduced as covariates to approximate unobserved heterogeneity and they then estimated car survival rates from a sample of 678 cars. The main result of their scrappage model suggests that vehicle age and total mileage are key determinants of scrappage decisions.
1.3. Goals

It is known that a survival approach to longevity obtains a better understanding of car ownership behavior, better estimations and assessments of GHG emissions, as well as finer estimations and forecasts of fleet composition in the absence of direct taxes. This paper proposes a framework already proven in the fields of biometrics and biology to compute and analyze robust survival rates. As in a biometric approach, the methodology is based on a sample of vehicles sent for scrapping in order to tackle the problem of right censored data. This closes a major gap in order to improve the current practice essentially based on datasets of active vehicles.

This paper further aims to create a survival-based tool for estimating robust scrapping rates of passenger cars that players can implement conveniently within a relatively short time in any country regardless of inadequacies in fleet data. We chose the survival approach because it promises detailed knowledge of all categories of road vehicles and thus sharpens the accuracy of GHG and other pollutant emissions forecasting. Our research applies the emissions factors and functions given in the European Environmental Agency’s (EEA) “Methodology for Calculating Transport Emissions and Energy Consumption” (MEET), which provides a Europe-wide standard to estimate GHG and other pollutant emissions and enables each country to estimate aggregated or disaggregated vehicle emissions and energy consumption.

In this paper, section 2 presents the dataset of scrapped vehicles along with the collected variables and underlying assumptions; sections 3 and 4 present the non-parametric and parametric results; section 5 and 6 show the same approaches based on total mileage, and section 6 addresses the issues of optimal choice for parametric survival models based on the standard parametric survival trio of lognormal, loglogistic and Weibull models. We then conclude in section 7.

2. DATA

Most studies estimate scrappage from smog check data (e.g. Yamamoto et al. 2004), vehicle registrations or annual census data, but we constructed a dataset exclusively for scrapped vehicles in order to consider and dissociate the effect of age deterioration from that of running. As given in Table 1 and Appendix, it comprises 265 vehicles from the following sources:

- Descriptions of 31 vehicles are drawn from a field survey at Georget Dépannage, an auto wrecker outside Paris.

- Data for 102 vehicles are from the European website Autoscout24 where cars are sold for spare parts.

- And finally, data for another 132 cars was lifted from Changeauto, a scrap dealers’ website in western France that trades in damaged cars.

---

4 See the Transport Research Knowledge Center at http://www.transport-research.info/web/projects/
5 http://www.autoscout24.com
6 http://www.changeauto.com
Data collected for each vehicle covered 11 key variables: make, model, version, vintage year, vintage month, type of fuel (mainly gasoline or diesel), year of scrappage, month of scrappage, vehicle condition (damaged or not), fiscal horsepower\(^7\) and odometer reading. Where the online sources omitted cylinder capacity and fiscal horsepower, we retrieved the missing data from Autoweb-france,\(^8\) which maintains the most comprehensive set of auto spec sheets available.

The date of scrapping, i.e. the exact date of the breakdown or accident that triggers the owner decision to scrap, almost always escapes precise determination. Scrappers opine there is a lag of one to four weeks between the moment of breakdown or accident and the actual date they take possession of the victimized vehicle.

Table 1 Vehicle Data by Source

<table>
<thead>
<tr>
<th>Sample</th>
<th>No. of vehicles</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Georget Depannage</td>
<td>31</td>
<td>11.7</td>
</tr>
<tr>
<td>Autoscout24</td>
<td>102</td>
<td>38.5</td>
</tr>
<tr>
<td>Changeauto</td>
<td>132</td>
<td>49.8</td>
</tr>
</tbody>
</table>

Damaged cars constitute 53.2 percent of vehicles in the database (Table 2) and the average damaged car was 92 months old (7y8m) with 100,835 km (63,022 miles) of mileage (Table 2). Meanwhile, 47.8 percent cars were scrapped because of mechanical breakdown or physical deterioration\(^9\) (Table 2) and here the average age and mileage are 220 months (18y4m) and 236,863 km (148,039 miles), over twice the figure for damaged vehicles. For all vehicles, the average age is 148 months (12y4m) and average mileage is 160,379 km (100,237 miles) (Table 2).

Table 2 Age and Mileage by Condition

<table>
<thead>
<tr>
<th>Vehicle Condition</th>
<th>Number</th>
<th>Percents</th>
<th>Average mileage in km (std-error)</th>
<th>Average age in months (std-error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Damaged cars</td>
<td>149</td>
<td>53.2</td>
<td>100,835 (62,620)</td>
<td>92 (43)</td>
</tr>
<tr>
<td>Non damaged cars</td>
<td>116</td>
<td>47.8</td>
<td>236,863 (85,515)</td>
<td>220 (64)</td>
</tr>
<tr>
<td>Total sample</td>
<td>265</td>
<td>100.0</td>
<td>160,379 (99,775)</td>
<td>148 (83)</td>
</tr>
</tbody>
</table>

\(^7\) In France, fiscal horsepower determine the amount of the registration card (paid either the car is new or second-hand). Since 1998, the formula to calculate the horsepower of a vehicle tax is:

\[
FH = \frac{CO_2}{45} + \left(\frac{H}{40}\right)^4
\]

where FH is the fiscal horsepower, \(CO_2\) is \(CO_2\) emissions in g/km and \(H\) the horsepower in kW.

\(^8\) [http://www.autoweb-france.com/fiches-techniques](http://www.autoweb-france.com/fiches-techniques)

\(^9\) Leading individual or multiple failures triggering the decision to scrap concerned failures of fuel or oil systems, cylinders, distributor, transmission, suspension, metal hose, contacts, piston rods, steering, clutch, catalyzer, gaskets, steering column, ignition, water pump.
3. Kaplan-Meier Non Parametric Analysis

Analysis of survival data usually begins by estimating the distribution of survival times. We borrowed Kaplan-Meier analysis (1958) from biometrics where it usefully calculates empirical life expectancy for human populations. From there, we input date of registration as the date of birth, thereby obtaining survival curves by following each car until death by scrappage and calculating survival at periodic waypoints. A Kaplan-Meier analysis enables estimation of survival trends over time, even where cars drop out or are studied for unequal periods of time. But, our case study tracked all cars through to scrappage and we have complete destruction data for almost the entire sample; thus right-censoring was not a consideration. For each interval, survival probability is calculated as the number of cars still in the fleet divided by the number of cars at risk. Cars that were scrapped, abandoned or reached their lifetime did not count anymore as 'at risk'. The probability of surviving to any point in time is estimated from the cumulative probability of surviving all preceding time intervals, i.e. survival is the product of all preceding probabilities. The resulting distribution gives the share of cars surviving at the start of each interval and equals the survival probabilities of preceding intervals multiplied by each other. $S(t)$ is the probability that any vehicle in the fleet will have a lifetime exceeding $t$. For a sample of a fleet of size $n$, the observed time to death is given by:

$$t_1 < t_2 < t_3 < \cdots < t_n$$

For each $t_i$, $m_i$ is the number at risk up to $t_i$ (risk-set) and $d_i$ is the number of deaths at $t_i$. Note that intervals are of unequal duration. The Kaplan-Meier estimator is the nonparametric maximum probability estimate of:

$$S_t = \prod_{t_j < t} (1 - \theta_j)$$

Is given by:

$$\tilde{S}_t = \prod_{t_j < t} (1 - \hat{\theta}_j) = \prod_{t_j < t} \left( \frac{r_j - d_j}{r_j} \right)$$

There is no censoring in our study and $n_i$ is the number of survivors just prior to $t_i$. With censoring, $n_i$ is the number of survivors less the number of losses, i.e. censored cases. Only the surviving, as yet uncensored, cases still being observed are "at risk".
4. Results of the Non Parametric Estimation for Age Dimension

We used a standard biometric approach for an overview of longevity with Kaplan-Meier survival curves for age and total mileage read from odometer readings. We then show that the non-parametric survival curves follow a two-parameter Weibull form.

4.1. Car Lifetime

Figure 1 depicts results of the non-parametric study of our sample. Kaplan-Meier survival analysis obtains median survival of 127 months (10y7m) with a 95% CI (120-131) 15 percent survive 240 months and some 3 percent approach 312 months.

4.2. Car Lifetime by Fuel Type

Average diesel car lifetime at scrappage is 121 months (10y1m) with corresponding total mileage of 170,394 km (106,496 miles) (Table 3). Meanwhile, gasoline cars average a higher 144 months (12y) in age at scrappage but with lower odometer readings which average out to 151,962 km (94,976 miles) (Table 3).

Our results show a substantial difference in age and mileage between gasoline and diesel cars. We began by comparing usage for each fuel type in order to measure any difference in life expectancy (Figure 2).

<table>
<thead>
<tr>
<th>Engine</th>
<th>Qty</th>
<th>Average mileage in km (std-error)</th>
<th>Average age in months (std-error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gasoline cars</td>
<td>144</td>
<td>151,962 (94,454)</td>
<td>163 (87)</td>
</tr>
<tr>
<td>Diesel cars</td>
<td>121</td>
<td>170,394 (78,174)</td>
<td>121 (75)</td>
</tr>
<tr>
<td>Total sample</td>
<td>265</td>
<td>160,379 (99,775)</td>
<td>148 (83)</td>
</tr>
</tbody>
</table>

Figure 2 shows life expectancy by fuel type. Gasoline cars have the best survival rate during the whole period, with 50 percent outliving 149 months (12y5m) (95% CI = 132-156), against 117 months (9y9m) (95% CI = 108-123) for diesel. Results are not surprising because we know that the diesel fleet experiences more intensive usage. This usage offsets the mechanical
and thermodynamic superiority of the diesel engine and seriously curtails life expectancy. Fuel type is therefore a major determinant of vehicle usage and exerts strong impact on vehicle longevity.

![Figure 2 - Kaplan-Meier age survival curves by fuel (in months)](image)

**5. Results of the Non-Parametric Mileage Estimation**

Unobserved heterogeneity is a major difficulty in building a semi-parametric survival model where cumulative mileage is a covariate. It may come from inclusion of time-sensitive variables included as covariates or from the omission of other important variables. Chen and Niemeier (2005), Chen and Lin (2006) and others treated unobserved heterogeneity by using a discrete approach. They followed Heckman and Singer (1985) and assumed the existence of two mass points. Unobserved heterogeneity can also be handled through a continuous approach that assumes a specific distribution, as did MacDonald (1984) or MacDonald and Butler (1987). Studies that incorporated odometer readings in a Cox proportional hazard model (Cox 1972) handle these data as a classic variable which requires special treatment to tackle unobserved heterogeneity from the model, e.g. Chen and Lin (2006), Lin et al. (2008).

Most researchers use age as the standard variable of lifetime measurement but others show that the natural time scale of the survival process is not calendar or clock time. Mathematical research on different time scales exists, including Oakes (1995) or Kordonsky and Gertsbakh (1997), who looked at multiple running time scales in survival data analysis. Duchesne and Lawless (2000) and Duchesne and Rosenthal (2003) described various advances in running time models for survival data while Cox and Oakes (1984) point out that "often the scale for measuring time is clock time, although other possibilities certainly arise, such as … mileage”.

For example, taxis and rented cars accumulate mileage far faster than privately owned cars and therefore odometer readings offer a better indication of longevity than age. This study underscores the advantage of using odometer readings as a dimension of design life rather than as a covariate.
5.1. Car Life and Total Mileage

Basing survival curves on the total mileage dimension is very convenient to determine the design life and vehicle reliability. Our results show that over 67 percent of vehicles exceed 100,000 km (62,500 miles) (95% CI = 91,000 – 105,000) and that median lifetime exceeds 150,000 km (94,000 miles) (95% CI = 141,530 – 105,000). Almost 10 percent approached 300,000 km (187,500 miles) (Figure 3).

![Figure 3 Kaplan-Meier mileage survival curves (in km)](image)

5.2. Fuel Types and Total Mileage

Figure 4 depicts results of the non-parametric study by fuel type. Diesel survival rates are higher than for gasoline in terms of total mileage. Kaplan-Meier survival analysis results shows that 50 percent of diesel vehicles survive beyond 152,730 km (95,456 miles) (95% CI = 149,686 – 163,411). The median lifetime of gasoline vehicles is 135,479 km (84,674 miles) (95% CI = 122,745 – 154,461). Results confirm that diesel vehicles have longer design life than gasoline vehicles despite the inverse relationship in terms of lifetime. Again, the relatively more intensive exploitation of diesel cars explains the inversion.
6. PARAMETRIC ANALYSIS

A parametric approach usefully compares non-parametric survival curves given for vehicles of different vintages by pinning down trends in survival function parameters. If we want a robust survival parametric model, we need to address the issue of comparative survival functions. Gallez (1994) used a lognormal survival function to fit French transversal scrapping rates estimated from panel data. However, the lognormal distribution also finds favor because of its convenience for transformation purposes, with results that are sufficiently accurate for fits and hypothesis testing but most studies use a Weibull survival function without alternatives (Zachariadis et al., 1995; Chen and Niemeier, 2004).

6.1. Description

If $T$ is a continuous random variable, we can define $F(t) = Pr(T < t)$ as the probability that the random variable $T$ is less than any value for $t$. The corresponding density function is given by $f(t) = df(t)/dt$. It is also useful to define the survival function $S(t) = 1 - F(t) = P(t \leq T)$, which is the probability that $T \geq t$. Also particularly useful for duration analysis is the hazard function defined as $h(t) = f(t)/S(t)$ which gives the probability of exit from the state immediately after $t$, given that the state is still occupied at $t$. Thus, if either $h(t), S(t), f(t)$ or $F(t)$ is known, the other values can be derived.
6.2. Parametric Survival Functions

This section presents lognormal, loglogistic and Weibull survival functions regularly used to fit vehicle (or human) survival rates before comparing them according to the criteria of the correlation coefficient and Fisher statistic on our Kaplan-Meier survival rates in months.

6.2.1 Lognormal Survival Function

The lognormal survival function is useful to represent lifetimes with increasing and decreasing hazard functions. It is expressed as:

\[
S(t) = \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left( \frac{\ln t - \ln m}{\sigma \sqrt{2}} \right)
\]

Where \( \operatorname{erf}(t) \) is the error function defined in:

\[
\operatorname{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-u^2} du
\]

6.2.2 Loglogistic Survival Function

The loglogistic survival function expresses lognormal lifetimes with increasing and decreasing hazard functions and is characterized by two constants \( \gamma > 0 \) and \( \alpha > 0 \):

\[
S(t) = \frac{1}{1 + \gamma t^\alpha}
\]

6.2.3 Weibull Survival Function

The Weibull distribution is a mainstay of life data analysis because of its versatility in fitting a variety of distributions. It furthermore offers a reasonable model for lifetimes of several types of units. The three-parameter Weibull pdf is defined as follows:

\[
f(t, \alpha, \beta, \theta) = \begin{cases} \\
\frac{\alpha}{\beta} \left( \frac{t - \theta}{\beta} \right)^{\alpha-1} e^{-(\frac{t-\theta}{\beta})^\alpha} & \text{if } t \geq \theta \\
0 & \text{if } t < \theta \\
\end{cases}
\]

where \( \alpha \) and \( \beta \) denote positive shape and scale parameters and \( \theta \), the location parameter (\( \theta = 0 \) yields to the two-parameter Weibull distribution. Note however that if \( \alpha = 1 \) we lapse into an exponential model. The \( \alpha \) is what gives this distribution its flexibility. The three- and two-parameter survival functions are given by:

\[
S(t, \alpha, \beta, \theta) = e^{-(\frac{t-\theta}{\beta})^\alpha} \quad \text{if } t \geq \theta \text{ and equal zero if } t < \theta
\]

\[
S(t, \alpha, \beta) = e^{-(\frac{t}{\beta})^\alpha}
\]
The $\beta$ parameter expresses the characteristic life of the Weibull distribution, which is the point at which the value of the survival function is $e^{-1}$, such that 63.2 percent of the fleet fails by $\beta$.

6.2.4. Optimal Choice

Figures 5.1 to 5.3 show our adjustments, as computed on the non-parametric survival rates for the entire sample by age in months using Weibull, lognormal and loglogistic survival. All three fit the data satisfactorily, with major deviations mainly in the tails of the functions. Nonetheless, our preliminary tests (Figures 5.1 to 5.3) show that the Weibull form obtains the best fit with real data at an average correlation coefficient of 0.998 and F statistic of 28898 (Table 4). The lognormal ($R^2 = 0.996; F = 11736$) is far behind the Weibull form but still outperforms the loglogistic ($R^2 = 0.994; F = 8451$). These results show that the Weibull survival function is well suited to model progressive car deterioration from mechanical fatigue.

<table>
<thead>
<tr>
<th>Survival function</th>
<th>$R^2$</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lognormal</td>
<td>.996</td>
<td>11736</td>
</tr>
<tr>
<td>Loglogistic</td>
<td>.994</td>
<td>8451</td>
</tr>
<tr>
<td>Weibull</td>
<td>.998</td>
<td>28893</td>
</tr>
</tbody>
</table>

The Newton-Raphson algorithm, having a quadratic rate of convergence, was used to optimize the likelihood equations which resolve the unknown parameters with maximum likelihood estimates.
6.3. Weibull Survival Function Parameter Estimation

In subsection 6.2.3 we showed that the Weibull survival function gives the best fit on computed survival rates by the Kaplan-Meier estimator. In this part of the paper, we made the choice to present all results fitted by a Weibull survival function. Table 5 lists the parameters of Weibull survival function given for vehicles age dimensions and Table 6 lists these parameters for mileage dimension.

6.3.1. Age Dimension

The Weibull survival function produces a very good fit with the non-parametric function by age for the entire sample of scrapped vehicles ($R^2 = 0.998$) as well as by fuel type ($R^2$ for diesel and gasoline are 0.984 and 0.993 respectively) (Table 5).

The Weibull shape parameter ($\alpha$) is 1.85, 1.79 and 1.99 for total, diesel and gasoline vehicle fleets respectively (Table 5). These results indicate a strong positive correlation between the design life and hazard functions. Meanwhile, the Weibull scale parameter $\beta$, or 63rd percentile of the survival function, is 166.5 for the entire sample. For diesel cars, $\beta$ is practically 145, far below 185 for gasoline cars.

![Figure 6 Weibull Survival Function Adjustment to Kaplan-Meier Survival Curve](image)
Figure 7 Weibull Survival Function Adjustment to Kaplan-Meier Survival Curve by Fuel Type

Table 5 Fit Parameters by Fuel for Age Dimension

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>95% confidence interval for $\alpha$</th>
<th>95% confidence interval for $\beta$</th>
<th>$R^2$</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gasoline</td>
<td>1.9895</td>
<td>184.67</td>
<td>[1.7525 ; 2.2586]</td>
<td>[169.35 ; 201.36]</td>
<td>.986</td>
<td>22249</td>
</tr>
<tr>
<td>Diesel</td>
<td>1.7884</td>
<td>145.13</td>
<td>[1.5509 ; 2.0622]</td>
<td>[130.68 ; 161.18]</td>
<td>.984</td>
<td>21137</td>
</tr>
<tr>
<td>Total sample</td>
<td>1.8528</td>
<td>166.46</td>
<td>[1.6857 ; 2.0365]</td>
<td>[155.45 ; 178.26]</td>
<td>.998</td>
<td>28893</td>
</tr>
</tbody>
</table>

6.3.2. Mileage Dimension

Mileage is treated as a dimension to measure design life and fundamental to any understanding of patterns in progressive vehicle deterioration. It also serves to distinguish all determinants involved in vehicle failure.

For mileage, the Weibull survival function secures a better fit with the non-parametric function ($R^2 = 0.998$) (Figure 8 and Table 6). For total the fleet, $\alpha$ approximates 1.616. The values for diesel and gasoline vehicles are 1.62 and 1.64 respectively (Table 6). These results indicate that the curvature of the survival law differentiates diesel from gasoline vehicles. Meanwhile, $\beta$ averages 178,717 km (111,698 miles), which breaks down into 189,612 km (118,508 miles) for diesel and 169,618 km (106,011 miles) gasoline vehicles. All else being equal, this shows that diesel vehicles have distinctly greater engineering reliability than their gasoline peers.
Table 6: Mileage Fit Parameters by Fuel Type

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>95% confidence interval for $\alpha$</th>
<th>95% confidence interval for $\beta$</th>
<th>$R^2$</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gasoline</td>
<td>1.6394</td>
<td>169618</td>
<td>[1.4414 ; 1.8646]</td>
<td>[152747 ; 188353]</td>
<td>.990</td>
<td>26592</td>
</tr>
<tr>
<td>Diesel</td>
<td>1.6159</td>
<td>189612</td>
<td>[1.3986 ; 1.8670]</td>
<td>[168888 ; 212879]</td>
<td>.985</td>
<td>22878</td>
</tr>
<tr>
<td>Total sample</td>
<td>1.6204</td>
<td>178719</td>
<td>[1.4720 ; 1.7838]</td>
<td>[165295 ; 193234]</td>
<td>.998</td>
<td>35776</td>
</tr>
</tbody>
</table>
7. CONCLUSION

This paper presents a new methodology to estimate vehicle lifetime based on field surveys of scrapped vehicles. This method is easy for government agencies, automakers and other players to implement within a relatively short timeframe in order to improve the accuracy of their fleet and GHG estimations, even if working with incomplete fleet data. Moreover, this study presents both parametric and non-parametric approaches to a sample of scrapped vehicles. This paper improves current practice by exploiting vehicle mileage as a dimension, rather than a determinant, of longevity in order to measure design life in addition to the standard dimension of age. Most studies use odometer readings as a covariate in semi-parametric survival models but vehicle longevity is likely a function of driver behavior which in turn is closely related to vehicle usage and maintenance frequency.

An automobile is a technologically advanced ensemble of complex components, each of which has its own design life and longevity characteristics. By increasing component design lives, automakers expect longer a service life from the vehicles they sell. However, vehicle longevity is mostly explained by driver behavior. This distinction between a vehicle’s intrinsic durability and its driver’s behavior is useful to understand the impact of various determinants on vehicle longevity.

Future research should focus on the following:

- Although our results show fuel type is a key determinant of vehicle longevity, we still need to measure the corresponding impact of determinants like vehicle rank, road types and vehicle price.
- We also need to distinguish between fuel type and cylinder capacity with the MEET methodology in order to estimate country emissions by estimating a survival function for six categories of passenger cars (Diesel<1.4L; 1.4L<Diesel<2L; Diesel>2L; Gasoline<1.4L; 1.4L< Gasoline <2L; Gasoline >2L).
- It would be useful to include damaged vehicles in the sample. Indeed, the share of damaged vehicles in the database is crucial to reliable estimations of service life. This would require collecting data on annual basis for 500 to 1,000 vehicles from various regions of France for input into a nationwide database of scrapped vehicles.
- The addition of household characteristics would provide another important breakthrough in the field of vehicle longevity.

**ACKNOWLEDGMENTS:** The author is grateful to Anu Tuominen from VVT (Technical Research Centre of Finland) for valuable suggestions which have significantly improved this paper and to Arthur Borges for editing. The author is responsible for all remaining errors.
REFERENCES


Weibull, W.: Composition And Decomposition Of Bounded Variates With Special Reference To The Gamma And The Weibull Distributions, USAF report AF61(052)-522; indexed as report AD-816012, (1967).
Transport sustainability and environment

19

APPENDIX: DATA
MAKE

MODEL

AUDI

80
CABRIOLET

CHRYSLER

VOYAGER

2.5 TD SE

PEUGEOT

306

1.9 XRD

RENAULT

SAFRANE

BACCARA
V6

FORD

PROBE

NISSAN

PRIMERA

LANCIA

KAPPA

VOLVO

S40

HYUNDAI

PONY

RENAULT

SAFRANE

RENAULT

CLIO II

OPEL

ASTRA G

RENAULT

CLIO

NISSAN

ALMEIRA

PEUGEOT

206

1.4HDI

OPEL

CORSA B

1.2i CITY

JAGUAR

S-TYPE

HONDA

CIVIC

RENAULT

ESPACE III

FORD

FOCUS

SUZUKI

WAGON R+

1.0 GL

PEUGEOT

206

1.4i 75CH
BVA

OPEL

CORSA

1.2i

RENAULT

TWINGO

PRIVILEGE

FORD

KA

1.3i ELANCE

DAEWOO

MATIZ

SE

RENAULT

CLIO

1.5 DCI
AUTHENTIQ
UE

PEUGEOT

BOXER

HDI 330 MH

RENAULT

VELSATIS

3.5 V6

FIAT

SEICENTO

SEAT

IBIZA III

PEUGEOT

206

PEUGEOT

206

VERSION

CYLINDER
CAPACITY

FINAL
ODOMETER
READING

YEAR OF
REGISTRATION

130 000

1993

7
6
16

237 685
297 111
95 000

1993
1994
1995

2200
1900

9
6
7
10
6
7
6

52 235
149 686
277 716
17 000
135 479
53 633
151 570

2000

5

1400
1200

(in cc)
2.3 E

2.0 TD SLX

2300

2.2 DT RXE

1.4i TS
AUTO
2.2 DT RXE
7PL

1.1i TEAM
1.9 TDI 100
FRESH
1.4 HDI
URBAN

1.6 HDI
1.6 HDI
CONFORT
172C
BENNE
DIESEL

PEUGEOT

PARTNER

PEUGEOT

504

VOLKSWAG
EN

GOLF I CAB

PEUGEOT

405

GRX TURBO
DIESEL

VOLVO

850

GLT 2.5

VOLVO
VOLKSWAG
EN
VOLKSWAG
EN

850

2.5 20V GLT

GOLF III
CAB

AVANTGAR
DE 90

POLO III

MATCH 3P

BUICK

PARK
AVENUE

CHRYSLER

VOYAGER

2.5L

FIAT

BRAVO

1.4 SX 12V

MERCEDES

E230

BREAK
AUTOMATIQ
UE

FUEL TYPE
1 =Gasoline
2=Diesel

MONTH OF
REGISTRATION

YEAR OF
SCRAPPAGE

MONTH OF
SCRAPPAGE

DAMMAGED
1=yes
0=no

1

2008

11

1

10
3
10

2
2
1

2008
2008
2008

12
12
9

0
0
0

1996
1997
1997
1998
1998
1998
1998

1
1
1
11
2
9
6

1
2
1
1
2
2

2008
2008
2008
2008
2008
2008
2008

9
12
9
9
10
9
12

1
1
1
1
1
0
1

110 000

1998

7

2

2008

10

1

1400

7
5
5
5
17
6

15 768
41 424
141 530
71 598
90 000
141 710

1998
1998
1999
1999
1999
1999

1
1
1
1
9
8

1
1
1
1

2008
2008
2008
2008
2008
2008

9
9
9
4
9
12

1
1
1
1
1
1

2200

7

204 559

1999

5

2

2008

10

1

1000
1400

5
6

91 543
122 068
79 129

1999
2000
2000

9
2
1

1
1

2008
2008
2008

9
10
10

1
1
1

1200

5
5
5
4
4

91 094
55 073
62 527
103 760
66 810

2000
2001
2001
2002
2002

3
7
10
3
1

1
1
1
1
2

2008
2008
2008
2008
2008

9
4
12
10
9

1
1
1
1
1

9

2002
2003
2003
2005

7
6
10

2
1
1
2

2008
2008
2008
2008

10
11
12
10

1
1
0
1

2500
1900

2000

LS
1.9D RTE
5P
2.0 DTI 16
CD

FISCAL
POWER

1300
1500

3500
1100
1900

4
6

142 009
90 000
15 000
39 919

1400

4

24 214

2006

5

2

2008

10

1

1600
1600

6
6

17 261
46 472

2006
2007

4
2

2
2

2008
2008

4
12

1
1

9

154 293

1989

9

2

2009

3

1

6

95 614

1992

9

1

2009

10

1

5

159 604

1992

5

2

2009

3

1

11
11
9

203 732
238 000
155 773

1993
1993
1994

7
8
7

1
1
1

2009
2009
2009

2
2
9

0
0
1

6

170 523

1995

3

1

2009

3

0

16

82 961

1995

7

1

2009

10

1

7
7
13

180 000
104 165
228 592

1996
1996
1996

2
7
8

1
1
1

2009
2009
2009

2
10
10

0
1
1

2500
2500

2500
1400
2300


| Brand     | Model  | Year | Engine | Power | Torque | Fuel Consumption | Year | Model
|-----------|--------|------|--------|-------|---------|------------------|------|-------
| VOLVO     | V70    | 2000 | 2.5T   | 216    | 330     | 12.5 L/100 km   | 2009 | DSR
| CITROEN  | C5     | 2000 | 1.6    | 120    | 160     | 8.8 L/100 km    | 2009 | DSR
| PEUGEOT  | 406    | 2000 | 1.6    | 116    | 150     | 6.9 L/100 km    | 2009 | DSR
| VOLKSWAGEN| POLO   | 2000 | 1.4    | 100    | 130     | 6.5 L/100 km    | 2009 | DSR
| CITROEN  | XANTIA | 2000 | 2.0L   | 140    | 185     | 10.2 L/100 km   | 2009 | DSR
| CITROEN  | XSARA  | 2000 | 1.8D   | 120    | 200     | 8.0 L/100 km    | 2009 | DSR
| FIAT     | SEICENTO| 2000 | 1.2    | 90     | 125     | 6.9 L/100 km    | 2009 | DSR
| RENAULT  | SCENIC | 2000 | 1.9D   | 125    | 240     | 9.1 L/100 km    | 2009 | DSR
| VOLVO    | S40    | 2000 | 1.9D   | 125    | 240     | 8.9 L/100 km    | 2009 | DSR
| AUDI     | TT     | 2000 | 2.8T   | 333    | 420     | 10.8 L/100 km   | 2009 | DSR
| FORD     | MONDEO | 2000 | 2.0L   | 140    | 200     | 8.9 L/100 km    | 2009 | DSR
| RENAULT  | CLIO   | 2000 | 1.2    | 85     | 110     | 6.7 L/100 km    | 2009 | DSR
| RENAULT  | KANGOO| 2000 | 1.5D   | 110    | 175     | 8.2 L/100 km    | 2009 | DSR
| MERCEDES | A170   | 2000 | 1.2    | 90     | 125     | 6.9 L/100 km    | 2009 | DSR
| PEUGEOT  | 406 III| 2000| 2.0L   | 140    | 240     | 9.1 L/100 km    | 2009 | DSR
| RENAULT  | LAGUNA | 2000 | 1.8D   | 125    | 240     | 9.1 L/100 km    | 2009 | DSR
| RENAULT  | MEGANE | 2000 | 1.9D   | 125    | 240     | 9.1 L/100 km    | 2009 | DSR
| VOLVO    | V40    | 2000 | 2.0L   | 140    | 240     | 9.1 L/100 km    | 2009 | DSR
| SUZUKI   | WAGON R+ | 2000 | 1.3    | 90     | 125     | 6.9 L/100 km    | 2009 | DSR
| SUZUKI   | WAGON R+| 2000 | 1.4    | 100    | 130     | 6.9 L/100 km    | 2009 | DSR
| AUDI     | S3     | 2000 | 1.8T   | 230    | 310     | 10.5 L/100 km   | 2009 | DSR
| DAEWOO   | MATIZ  | 2000 | 1.0    | 75     | 100     | 5.9 L/100 km    | 2009 | DSR
| MITSUBISHI| CARISMA| 2000| 1.3    | 90     | 125     | 6.9 L/100 km    | 2009 | DSR
| PEUGEOT  | 206    | 2000 | 1.6T   | 200    | 260     | 9.1 L/100 km    | 2009 | DSR
| PEUGEOT  | 306    | 2000 | 1.8T   | 230    | 310     | 10.5 L/100 km   | 2009 | DSR
| FORD     | MONDEO | 2000 | 2.0L   | 140    | 200     | 8.9 L/100 km    | 2009 | DSR
| RENAULT  | CLIO   | 2000 | 1.2    | 85     | 110     | 6.7 L/100 km    | 2009 | DSR
| RENAULT  | TWINGO| 2000 | 1.2L   | 85     | 110     | 6.7 L/100 km    | 2009 | DSR
| VOLKSWAGEN| POLO IV| 2000| TDI 3P | 110    | 150     | 7.2 L/100 km    | 2009 | DSR
| OPEL     | MOVANO | 2000 | 2.0L   | 140    | 200     | 8.9 L/100 km    | 2009 | DSR
| PEUGEOT  | 306    | 2000 | 2.0L   | 140    | 200     | 8.9 L/100 km    | 2009 | DSR
| RENAULT  | CLIO   | 2000 | 1.2    | 85     | 110     | 6.7 L/100 km    | 2009 | DSR
| RENAULT  | MEGANE | 2000 | 1.2    | 85     | 110     | 6.7 L/100 km    | 2009 | DSR
| RENAULT  | KANGOO| 2000 | 1.9D   | 125    | 240     | 9.1 L/100 km    | 2009 | DSR
| CITROEN  | SAXO   | 2000 | 1.6D   | 120    | 160     | 7.9 L/100 km    | 2009 | DSR
| CITROEN  | PICASSO| 2000| 1.6L   | 100    | 130     | 7.2 L/100 km    | 2009 | DSR
| FORD     | FIESTA IV| 2000| 1.4L   | 90     | 125     | 6.9 L/100 km    | 2009 | DSR
| FORD     | FIESTA IV| 2000| 1.2L   | 85     | 110     | 6.7 L/100 km    | 2009 | DSR
| PEUGEOT  | 206    | 2000 | 2.0L   | 140    | 200     | 8.9 L/100 km    | 2009 | DSR
| PEUGEOT  | 307    | 2000 | 1.6L   | 120    | 160     | 9.1 L/100 km    | 2009 | DSR
| PEUGEOT  | 206    | 2000 | 1.8L   | 140    | 200     | 9.1 L/100 km    | 2009 | DSR
| RENAULT  | CLIO   | 2000 | 1.4L   | 100    | 130     | 7.2 L/100 km    | 2009 | DSR
| RENAULT  | TWINGO| 2000 | 1.2L   | 85     | 110     | 6.7 L/100 km    | 2009 | DSR
| VOLVO    | S40    | 2000 | 1.9D   | 125    | 240     | 9.1 L/100 km    | 2009 | DSR
<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Model</th>
<th>Year</th>
<th>Engine Capacity</th>
<th>Transmission</th>
<th>Year</th>
<th>Fuel Efficiency</th>
<th>CO2 Emissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>VOLVO</td>
<td>V70</td>
<td>2002</td>
<td>2.5L</td>
<td>5-speed</td>
<td>2009</td>
<td>115,150</td>
<td>2</td>
</tr>
<tr>
<td>CITROEN</td>
<td>XSARA II</td>
<td>1.9D SX 3P</td>
<td>1990</td>
<td>5</td>
<td>105,546</td>
<td>2009</td>
<td>11</td>
</tr>
<tr>
<td>CITROEN</td>
<td>XSARA II</td>
<td>2.0 HDI</td>
<td>2000</td>
<td>5</td>
<td>227,139</td>
<td>2009</td>
<td>10</td>
</tr>
<tr>
<td>CITROEN</td>
<td>XSARA II</td>
<td>1.9D SX 3P</td>
<td>1990</td>
<td>5</td>
<td>141,906</td>
<td>2009</td>
<td>6</td>
</tr>
<tr>
<td>CITROEN</td>
<td>SAXO II</td>
<td>1.9D EXCLUSIVE SP</td>
<td>1500</td>
<td>4</td>
<td>114,310</td>
<td>2009</td>
<td>2</td>
</tr>
<tr>
<td>CITROEN</td>
<td>BERLINGO</td>
<td>2.0 HDI 800KG FOURGON</td>
<td>2000</td>
<td>8</td>
<td>123,321</td>
<td>2009</td>
<td>10</td>
</tr>
<tr>
<td>FIAT</td>
<td>PUNTO II</td>
<td>JTD ELX 3P</td>
<td>5</td>
<td>94,000</td>
<td>2009</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>306</td>
<td>1.6L V6</td>
<td>1800</td>
<td>7</td>
<td>108,420</td>
<td>2009</td>
<td>10</td>
</tr>
<tr>
<td>FIAT</td>
<td>PUNTO II</td>
<td>80 DYNAMIC SP</td>
<td>60,706</td>
<td>2009</td>
<td>10</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>206</td>
<td>1.4 X LINE</td>
<td>1400</td>
<td>5</td>
<td>77,226</td>
<td>2009</td>
<td>10</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>206</td>
<td>1.4 AUTOMATIC SP</td>
<td>1400</td>
<td>5</td>
<td>40,000</td>
<td>2009</td>
<td>10</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>206</td>
<td>1.6 V5 S3</td>
<td>1600</td>
<td>7</td>
<td>102,825</td>
<td>2009</td>
<td>10</td>
</tr>
<tr>
<td>RENAULT</td>
<td>VELSATIS</td>
<td>3.0 V6 AUTO</td>
<td>3000</td>
<td>12</td>
<td>103,984</td>
<td>2009</td>
<td>9</td>
</tr>
<tr>
<td>RENAULT</td>
<td>TIWINGO</td>
<td>1.2 16V</td>
<td>1200</td>
<td>5</td>
<td>74,131</td>
<td>2009</td>
<td>9</td>
</tr>
<tr>
<td>CITROEN</td>
<td>SAXO II</td>
<td>1.5D</td>
<td>1500</td>
<td>6</td>
<td>51,753</td>
<td>2009</td>
<td>9</td>
</tr>
<tr>
<td>CITROEN</td>
<td>XSARA</td>
<td>2.0 HDI VTS</td>
<td>2000</td>
<td>5</td>
<td>157,669</td>
<td>2009</td>
<td>10</td>
</tr>
<tr>
<td>FORD</td>
<td>FUSION</td>
<td>1.4 TDCI TRENDS</td>
<td>1400</td>
<td>5</td>
<td>151,399</td>
<td>2009</td>
<td>9</td>
</tr>
<tr>
<td>OPEL</td>
<td>ASTRA</td>
<td>2.0 D TCi 16V</td>
<td>2000</td>
<td>6</td>
<td>150,610</td>
<td>2009</td>
<td>2</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>PARTNER</td>
<td>1.9D CONFORT 170C</td>
<td>1900</td>
<td>7</td>
<td>104,655</td>
<td>2009</td>
<td>10</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>206</td>
<td>1.8D CLIM 3P</td>
<td>1900</td>
<td>7</td>
<td>168,159</td>
<td>2009</td>
<td>5</td>
</tr>
<tr>
<td>RENAULT</td>
<td>KANGOO</td>
<td>1.9D EXPRESS CONFORT 3S</td>
<td>1900</td>
<td>7</td>
<td>84,268</td>
<td>2009</td>
<td>9</td>
</tr>
<tr>
<td>RENAULT</td>
<td>KANGOO</td>
<td>1.3DTI 80 BREAK</td>
<td>1900</td>
<td>5</td>
<td>165,870</td>
<td>2009</td>
<td>10</td>
</tr>
<tr>
<td>CITROEN</td>
<td>XSARA PICASSO</td>
<td>65,783</td>
<td>2009</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYUNDAI</td>
<td>MATRIX</td>
<td>CRD</td>
<td>85,000</td>
<td>2003</td>
<td>2</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>KIA</td>
<td>CARNIVAL</td>
<td>2.0 CRDI EX JUCKE</td>
<td>2900</td>
<td>10</td>
<td>114,763</td>
<td>2003</td>
<td>12</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>307</td>
<td>2.0 HDI 110 X 3P</td>
<td>2000</td>
<td>6</td>
<td>43,759</td>
<td>2003</td>
<td>4</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>206</td>
<td>2.0 HDI 110 X 3P</td>
<td>2000</td>
<td>5</td>
<td>152,316</td>
<td>2003</td>
<td>7</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>206</td>
<td>2.0 HDI</td>
<td>2000</td>
<td>5</td>
<td>125,431</td>
<td>2003</td>
<td>10</td>
</tr>
<tr>
<td>ALFA ROMEO</td>
<td>GTV SPIDER</td>
<td>2.0 JTS DISTINCTIV E</td>
<td>2000</td>
<td>11</td>
<td>43,538</td>
<td>2004</td>
<td>10</td>
</tr>
<tr>
<td>CITROEN</td>
<td>XSARA II</td>
<td>1.6L SX 3P</td>
<td>1600</td>
<td>7</td>
<td>87,963</td>
<td>2004</td>
<td>10</td>
</tr>
<tr>
<td>FIAT</td>
<td>PUNTO II</td>
<td>1.3L AMBIENTE 3P</td>
<td>1400</td>
<td>5</td>
<td>50,550</td>
<td>2005</td>
<td>10</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>206</td>
<td>1.4 T XLINE</td>
<td>1400</td>
<td>5</td>
<td>39,940</td>
<td>2004</td>
<td>10</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>206</td>
<td>1.4 JTD</td>
<td>1400</td>
<td>5</td>
<td>47,375</td>
<td>2004</td>
<td>12</td>
</tr>
<tr>
<td>CITROEN</td>
<td>BERLINGO</td>
<td>2.0 HDI</td>
<td>2000</td>
<td>8</td>
<td>49,547</td>
<td>2004</td>
<td>2</td>
</tr>
<tr>
<td>MG</td>
<td>ZS</td>
<td>1.5</td>
<td>115</td>
<td>1</td>
<td>117,752</td>
<td>2004</td>
<td>9</td>
</tr>
<tr>
<td>FORD</td>
<td>FESTA IV</td>
<td>1.3 AMBIENTE 3P</td>
<td>1300</td>
<td>5</td>
<td>50,550</td>
<td>2005</td>
<td>10</td>
</tr>
<tr>
<td>LANCIA</td>
<td>YPSILON</td>
<td>1.6L V6 ORO 3P</td>
<td>1200</td>
<td>5</td>
<td>44,722</td>
<td>2005</td>
<td>10</td>
</tr>
<tr>
<td>FORD</td>
<td>FESTA V</td>
<td>1.3TDCI 3P</td>
<td>4</td>
<td>112,806</td>
<td>2005</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>BOXER</td>
<td>2.2 HDI</td>
<td>2200</td>
<td>7</td>
<td>102,515</td>
<td>2005</td>
<td>10</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>307</td>
<td>1.9 TDI CLIM CD SOCIETE 1.5uth CONfort</td>
<td>2000</td>
<td>8</td>
<td>70,984</td>
<td>2005</td>
<td>5</td>
</tr>
<tr>
<td>RENAULT</td>
<td>KANGOO</td>
<td>1500</td>
<td>6</td>
<td>109,134</td>
<td>2005</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>RENAULT</td>
<td>CLIO II</td>
<td>1.5 D GTS AUTHENTIQ U</td>
<td>1500</td>
<td>4</td>
<td>75,264</td>
<td>2005</td>
<td>10</td>
</tr>
<tr>
<td>Car Make</td>
<td>Model</td>
<td>Engine Type</td>
<td>Year</td>
<td>Mileage</td>
<td>Age (years)</td>
<td>Odometer (miles)</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>-------------</td>
<td>------</td>
<td>---------</td>
<td>-------------</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td>RENAULT</td>
<td>CLIO II</td>
<td>1.5 DCI 65</td>
<td>2010</td>
<td>2</td>
<td>2009</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>VOLKSWAGEN</td>
<td>POLO V</td>
<td>1.4 TSI</td>
<td>2010</td>
<td>2</td>
<td>2009</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>FORD</td>
<td>FIESTA V</td>
<td>1.4 BENZ</td>
<td>2010</td>
<td>2</td>
<td>2009</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>207</td>
<td>1.4L 16V</td>
<td>2010</td>
<td>2</td>
<td>2009</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>1007</td>
<td>1.4DOLCE</td>
<td>2010</td>
<td>2</td>
<td>2009</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>MAZDA</td>
<td>PICK UP</td>
<td>2.5SAI</td>
<td>2010</td>
<td>2</td>
<td>2009</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>PARTNER</td>
<td>1.6 HDI</td>
<td>2010</td>
<td>2</td>
<td>2009</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>PARTNER</td>
<td>1.6 HDI</td>
<td>2010</td>
<td>2</td>
<td>2009</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>206</td>
<td>1.4L 16V</td>
<td>2010</td>
<td>2</td>
<td>2009</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>PARTNER</td>
<td>1.6 HDI</td>
<td>2010</td>
<td>2</td>
<td>2009</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>AUDI</td>
<td>A6</td>
<td>2.5TDI V6</td>
<td>2010</td>
<td>2</td>
<td>2009</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>FORD</td>
<td>PROBE</td>
<td>1.4L 16V</td>
<td>2010</td>
<td>2</td>
<td>2009</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>OPEL</td>
<td>CORSA</td>
<td>1.6L 16V</td>
<td>2010</td>
<td>2</td>
<td>2009</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>HONDA</td>
<td>ACURA</td>
<td>1.6L 16V</td>
<td>2010</td>
<td>2</td>
<td>2009</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>MERCEDES</td>
<td>E300</td>
<td>1.4L 16V</td>
<td>2010</td>
<td>2</td>
<td>2009</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Transport sustainability and environment
<table>
<thead>
<tr>
<th>Car Make</th>
<th>Model</th>
<th>Year</th>
<th>Engine Size (cc)</th>
<th>Horsepower</th>
<th>Torque (Nm)</th>
<th>Production Run (Units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPEL</td>
<td>CORSA</td>
<td>2003</td>
<td>1.4D GL</td>
<td>1488</td>
<td>5</td>
<td>240 000</td>
</tr>
<tr>
<td>Datsun</td>
<td>Oldtimer</td>
<td>1982</td>
<td>Cherry Deville 1.0</td>
<td>1000</td>
<td>5</td>
<td>206 302</td>
</tr>
<tr>
<td>Renault</td>
<td>R11</td>
<td>1983</td>
<td>GTX</td>
<td>1721</td>
<td>7</td>
<td>350 000</td>
</tr>
<tr>
<td>Renault</td>
<td>Super 5</td>
<td>1991</td>
<td>FIVE</td>
<td>1100</td>
<td>4</td>
<td>160 000</td>
</tr>
<tr>
<td>Ford</td>
<td>Sierra</td>
<td>1988</td>
<td>2.3L MK</td>
<td>2300</td>
<td>7</td>
<td>290 000</td>
</tr>
<tr>
<td>Peugeot</td>
<td>205</td>
<td>1984</td>
<td>GTi</td>
<td>1905</td>
<td>10</td>
<td>212 500</td>
</tr>
<tr>
<td>Volvo</td>
<td>760</td>
<td>1985</td>
<td>GLD TURBO</td>
<td>2316</td>
<td>8</td>
<td>353 393</td>
</tr>
<tr>
<td>Volkswagen</td>
<td>Golf II</td>
<td>1991</td>
<td>Boston</td>
<td>1588</td>
<td>7</td>
<td>126 000</td>
</tr>
<tr>
<td>Ford</td>
<td>Fiesta</td>
<td>1990</td>
<td>1.1</td>
<td>1118</td>
<td>5</td>
<td>250 000</td>
</tr>
<tr>
<td>Peugeot</td>
<td>205</td>
<td>1998</td>
<td>GENERATION</td>
<td>1769</td>
<td>5</td>
<td>250 000</td>
</tr>
<tr>
<td>Renault</td>
<td>R21</td>
<td>1992</td>
<td>TD</td>
<td>2068</td>
<td>8</td>
<td>250 000</td>
</tr>
<tr>
<td>OPEL</td>
<td>CORSA</td>
<td>1990</td>
<td>1.4S</td>
<td>1396</td>
<td>5</td>
<td>200 000</td>
</tr>
<tr>
<td>Peugeot</td>
<td>106</td>
<td>1994</td>
<td>1.4i</td>
<td>1360</td>
<td>5</td>
<td>295 000</td>
</tr>
<tr>
<td>Citroen</td>
<td>BX</td>
<td>1990</td>
<td>TZX</td>
<td>1905</td>
<td>5</td>
<td>300 000</td>
</tr>
<tr>
<td>Citroen</td>
<td>C15</td>
<td>1987</td>
<td>VOPA</td>
<td>1124</td>
<td>8</td>
<td>300 000</td>
</tr>
<tr>
<td>Volkswagen</td>
<td>Scirocco</td>
<td>1989</td>
<td>MEMPHIS</td>
<td>1800</td>
<td>7</td>
<td>213 587</td>
</tr>
<tr>
<td>Lancia</td>
<td>Y10</td>
<td>1993</td>
<td>1.1</td>
<td>1100</td>
<td>4</td>
<td>150 000</td>
</tr>
<tr>
<td>Peugeot</td>
<td>205</td>
<td>1983</td>
<td>1.1 GR</td>
<td>1124</td>
<td>4</td>
<td>302 000</td>
</tr>
<tr>
<td>Renault</td>
<td>Clio</td>
<td>1996</td>
<td>RN</td>
<td>1400</td>
<td>7</td>
<td>245 000</td>
</tr>
<tr>
<td>OPEL</td>
<td>ASTRA</td>
<td>1993</td>
<td>1.7D</td>
<td>1700</td>
<td>5</td>
<td>330 000</td>
</tr>
<tr>
<td>Volvo</td>
<td>480</td>
<td>1989</td>
<td>ES</td>
<td>1998</td>
<td>8</td>
<td>192 000</td>
</tr>
<tr>
<td>Peugeot</td>
<td>106</td>
<td>1996</td>
<td>OPEN 1.4L</td>
<td>1400</td>
<td>5</td>
<td>200 000</td>
</tr>
<tr>
<td>Renault</td>
<td>R5</td>
<td>1993</td>
<td>TR</td>
<td>1237</td>
<td>5</td>
<td>150 000</td>
</tr>
<tr>
<td>OPEL</td>
<td>CORSA</td>
<td>1997</td>
<td>CITY</td>
<td>1000</td>
<td>5</td>
<td>190 000</td>
</tr>
<tr>
<td>Ford</td>
<td>Fiesta</td>
<td>1992</td>
<td>1.3</td>
<td>1299</td>
<td>6</td>
<td>101 000</td>
</tr>
<tr>
<td>Ford</td>
<td>Escort</td>
<td>1991</td>
<td>1.8D</td>
<td>1753</td>
<td>5</td>
<td>210 000</td>
</tr>
<tr>
<td>Volkswagen</td>
<td>Golf</td>
<td>1996</td>
<td>1.8</td>
<td>1781</td>
<td>7</td>
<td>195 000</td>
</tr>
<tr>
<td>Ford</td>
<td>Fiesta</td>
<td>1997</td>
<td>1.8D</td>
<td>1753</td>
<td>6</td>
<td>250 000</td>
</tr>
<tr>
<td>Citroen</td>
<td>BX</td>
<td>1992</td>
<td>GTI</td>
<td>1905</td>
<td>9</td>
<td>155 000</td>
</tr>
<tr>
<td>Renault</td>
<td>R25</td>
<td>1987</td>
<td>TX</td>
<td>1995</td>
<td>8</td>
<td>200 500</td>
</tr>
<tr>
<td>OPEL</td>
<td>CORSA</td>
<td>1989</td>
<td>1.3 S</td>
<td>1297</td>
<td>5</td>
<td>189 000</td>
</tr>
<tr>
<td>BMW</td>
<td>SERIE 5</td>
<td>1989</td>
<td>524D</td>
<td>2400</td>
<td>7</td>
<td>300 000</td>
</tr>
<tr>
<td>DAEWOO</td>
<td>Matiz</td>
<td>1999</td>
<td>SE</td>
<td>995</td>
<td>4</td>
<td>299 000</td>
</tr>
<tr>
<td>Dodge</td>
<td>Grand Caravan</td>
<td>1990</td>
<td>SE V6</td>
<td>2450</td>
<td>16</td>
<td>179 000</td>
</tr>
<tr>
<td>SEAT</td>
<td>Cordoba</td>
<td>1994</td>
<td>GLX 1.9D</td>
<td>1896</td>
<td>7</td>
<td>180 000</td>
</tr>
<tr>
<td>Ford</td>
<td>Fiesta</td>
<td>1989</td>
<td>1.8D</td>
<td>1753</td>
<td>5</td>
<td>200 000</td>
</tr>
<tr>
<td>Renault</td>
<td>Megane</td>
<td>1996</td>
<td>1.9D</td>
<td>1870</td>
<td>7</td>
<td>270 000</td>
</tr>
<tr>
<td>CITROEN</td>
<td>BX</td>
<td>1986</td>
<td>UTAH</td>
<td>1000</td>
<td>5</td>
<td>250 000</td>
</tr>
<tr>
<td>Renault</td>
<td>R21</td>
<td>1992</td>
<td>GTS</td>
<td>2172</td>
<td>7</td>
<td>210 580</td>
</tr>
<tr>
<td>Renault</td>
<td>R25</td>
<td>1991</td>
<td>CVVL</td>
<td>2068</td>
<td>10</td>
<td>180 000</td>
</tr>
<tr>
<td>Rover</td>
<td>418</td>
<td>1992</td>
<td>GSD TURBO</td>
<td>1800</td>
<td>5</td>
<td>250 000</td>
</tr>
<tr>
<td>BMW</td>
<td>SERIE 3</td>
<td>1987</td>
<td>324D</td>
<td>2500</td>
<td>8</td>
<td>417 500</td>
</tr>
<tr>
<td>Volkswagen</td>
<td>Jetta</td>
<td>1982</td>
<td>DIESEL</td>
<td>2000</td>
<td>5</td>
<td>175 000</td>
</tr>
<tr>
<td>Renault</td>
<td>R25</td>
<td>1986</td>
<td>GTS</td>
<td>1995</td>
<td>9</td>
<td>285 000</td>
</tr>
<tr>
<td>CITROEN</td>
<td>Xantia</td>
<td>1998</td>
<td>TURBO EXCLUSIVE</td>
<td>1905</td>
<td>7</td>
<td>325 000</td>
</tr>
<tr>
<td>Hyundai</td>
<td>Lantra</td>
<td>1997</td>
<td>BREAK 1.6 GSi</td>
<td>1599</td>
<td>7</td>
<td>350 000</td>
</tr>
<tr>
<td>Peugeot</td>
<td>309</td>
<td>1990</td>
<td>GREEN</td>
<td>1400</td>
<td>7</td>
<td>167 000</td>
</tr>
<tr>
<td>Peugeot</td>
<td>205</td>
<td>1990</td>
<td>XA</td>
<td>1124</td>
<td>5</td>
<td>260 000</td>
</tr>
<tr>
<td>Citroen</td>
<td>ZX</td>
<td>1993</td>
<td>1.1i</td>
<td>1124</td>
<td>5</td>
<td>145 000</td>
</tr>
<tr>
<td>Renault</td>
<td>R21</td>
<td>1989</td>
<td>2.1TD</td>
<td>2100</td>
<td>9</td>
<td>424 000</td>
</tr>
<tr>
<td>Citroen</td>
<td>AX</td>
<td>1994</td>
<td>AUDACE</td>
<td>954</td>
<td>4</td>
<td>214 000</td>
</tr>
<tr>
<td>Peugeot</td>
<td>309</td>
<td>1990</td>
<td>GR</td>
<td>1400</td>
<td>7</td>
<td>221 000</td>
</tr>
<tr>
<td>Seat</td>
<td>Ibiza</td>
<td>1996</td>
<td>1.4</td>
<td>1380</td>
<td>5</td>
<td>182 000</td>
</tr>
<tr>
<td>Make</td>
<td>Model</td>
<td>Year</td>
<td>Age</td>
<td>Value</td>
<td>Mileage</td>
<td>Model Year</td>
</tr>
<tr>
<td>--------------</td>
<td>----------</td>
<td>------</td>
<td>-----</td>
<td>-------</td>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>HYUNDAI</td>
<td>LANTRA</td>
<td>1995</td>
<td>8</td>
<td>170 000</td>
<td>1997</td>
<td>4</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>205</td>
<td>1995</td>
<td>10</td>
<td>334 000</td>
<td>1991</td>
<td>1</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>206</td>
<td>1995</td>
<td>5</td>
<td>495 000</td>
<td>1995</td>
<td>8</td>
</tr>
<tr>
<td>RENAULT</td>
<td>R25</td>
<td>1995</td>
<td>4</td>
<td>206 000</td>
<td>1992</td>
<td>12</td>
</tr>
<tr>
<td>RENAULT</td>
<td>SUPER 5</td>
<td>1995</td>
<td>7</td>
<td>300 000</td>
<td>1997</td>
<td>11</td>
</tr>
<tr>
<td>FIAT</td>
<td>FIAT 1.6</td>
<td>1995</td>
<td>2</td>
<td>215 768</td>
<td>1989</td>
<td>2</td>
</tr>
<tr>
<td>RENAULT</td>
<td>R5</td>
<td>1995</td>
<td>6</td>
<td>450 000</td>
<td>1984</td>
<td>7</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>309</td>
<td>1995</td>
<td>11</td>
<td>260 000</td>
<td>1992</td>
<td>6</td>
</tr>
<tr>
<td>RENAULT</td>
<td>SUPER 5</td>
<td>1995</td>
<td>2</td>
<td>215 260</td>
<td>1995</td>
<td>2</td>
</tr>
<tr>
<td>CHRYSLER</td>
<td>VOYAGER</td>
<td>1995</td>
<td>8</td>
<td>198 563</td>
<td>1995</td>
<td>8</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>405</td>
<td>1995</td>
<td>7</td>
<td>250 000</td>
<td>1990</td>
<td>3</td>
</tr>
<tr>
<td>CITROEN</td>
<td>XANTIA</td>
<td>1995</td>
<td>7</td>
<td>361 000</td>
<td>1997</td>
<td>9</td>
</tr>
<tr>
<td>SEAT</td>
<td>IBIZA</td>
<td>1995</td>
<td>7</td>
<td>286 000</td>
<td>1994</td>
<td>9</td>
</tr>
<tr>
<td>RENAULT</td>
<td>LAGUNA</td>
<td>1995</td>
<td>7</td>
<td>286 000</td>
<td>1994</td>
<td>9</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>204</td>
<td>1995</td>
<td>7</td>
<td>271 000</td>
<td>1993</td>
<td>6</td>
</tr>
<tr>
<td>RENAULT</td>
<td>ESPACE</td>
<td>1995</td>
<td>7</td>
<td>335 000</td>
<td>1989</td>
<td>5</td>
</tr>
<tr>
<td>SEAT</td>
<td>TOLEDO</td>
<td>1995</td>
<td>6</td>
<td>280 000</td>
<td>1995</td>
<td>10</td>
</tr>
<tr>
<td>RENAULT</td>
<td>R19</td>
<td>1995</td>
<td>7</td>
<td>295 000</td>
<td>1993</td>
<td>4</td>
</tr>
<tr>
<td>RENAULT</td>
<td>R19</td>
<td>1995</td>
<td>6</td>
<td>160 000</td>
<td>1985</td>
<td>1</td>
</tr>
<tr>
<td>ROVER</td>
<td>213</td>
<td>1995</td>
<td>4</td>
<td>265 000</td>
<td>1989</td>
<td>11</td>
</tr>
<tr>
<td>RENAULT</td>
<td>SUPER 5</td>
<td>1995</td>
<td>5</td>
<td>280 000</td>
<td>1992</td>
<td>12</td>
</tr>
<tr>
<td>RENAULT</td>
<td>R21</td>
<td>1995</td>
<td>7</td>
<td>174 000</td>
<td>1988</td>
<td>2</td>
</tr>
<tr>
<td>VOLKSWAGEN</td>
<td>GOLF II</td>
<td>1995</td>
<td>5</td>
<td>200 000</td>
<td>1990</td>
<td>12</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>405</td>
<td>1995</td>
<td>5</td>
<td>140 000</td>
<td>1989</td>
<td>11</td>
</tr>
<tr>
<td>RENAULT</td>
<td>R19</td>
<td>1995</td>
<td>6</td>
<td>302 000</td>
<td>1992</td>
<td>10</td>
</tr>
<tr>
<td>VOLKSWAGEN</td>
<td>PASSAT III</td>
<td>1995</td>
<td>7</td>
<td>320 000</td>
<td>1993</td>
<td>2</td>
</tr>
<tr>
<td>RENAULT</td>
<td>R21</td>
<td>1995</td>
<td>7</td>
<td>300 000</td>
<td>1993</td>
<td>5</td>
</tr>
<tr>
<td>PEUGEOT</td>
<td>309</td>
<td>1995</td>
<td>7</td>
<td>130 000</td>
<td>1989</td>
<td>1</td>
</tr>
<tr>
<td>AUDI</td>
<td>CLIO</td>
<td>1995</td>
<td>6</td>
<td>200 000</td>
<td>1995</td>
<td>7</td>
</tr>
</tbody>
</table>
Impact of the penetration rate of ecodriving on fuel consumption and traffic congestion

Keywords: Ecodriving, Fuel consumption, Traffic congestion

Abstract

Since many years, as a consequence of fossil fuels rarefaction and climate changes, ecology has become a major challenge of our society. In this field, many improvements could be realized on the transportation side and more precisely on passenger cars which are an important source of pollution. A quick and efficient solution to reduce fuel consumption and so, greenhouse gases emissions, is to adopt an ecological way of driving, called ecodriving. However, is ecodriving really efficient in terms of mobility and environment at a global point of view? The benefits of ecodriving have often been studied for an isolated vehicle and rarely for a whole network. The aim of this work is to estimate the effects of ecodriving on traffic congestion and fuel consumption according to the percentage of ecodrivers in the population. This has been achieved using a class of ecodriven vehicles with a car-following model (Intelligent Driver Model) and with a transport simulation software (Aimsun). Results show that the effect of ecodriving on the traffic congestion and pollution is not linearly linked to the proportion of ecodrivers and this effect varies according to the driving conditions. In some cases, ecodriving is cons-productive and fuel consumption increases. Future works will concentrate on experimental validation, on modeling the effect of ecodriving on road safety and on improving the different models.
1. INTRODUCTION

1.1 Context

The fact that fossils fuels are running out faster than they are generated by nature is in all minds. While waiting for new and viable solutions to produce energy, the world population has to learn how to make the remaining reserves last as long as possible by reducing the consumption. More particularly, strong efforts should be concentrated on the transportation side as it represents 30% of our energy consumption (U. S. Energy Information Administration). Research is underway to improve the vehicle and road technologies to lower fuel consumption but a direct method, known as eco-driving, consisting in modifying the driving style, could be deployed rapidly. Although researchers are agree on the fact that ecodriving should reduce the fuel consumption for an isolated vehicle, the information about a whole population of drivers practicing ecodriving is still lacking. This information could be important for policy makers to assist them in the decision to promote or not ecodriving but also for car manufacturers to design driver assistance systems and for road managers to optimize their networks. At first, a precise definition of ecodriving is needed.

1.1.1. Ecodriving. In the literature, ecodriving is often referred as a way of driving allowing to reduce fuel consumption and so greenhouse gases emissions (Beusen et al., 2009; Barth and Boriboonsomsin, 2009). This is rather subjective and a precise definition could be difficult to propose. In most publications (Saboohi and Farzaneh, 2009; Wåhlberg, 2007; Zarkadoula et al., 2007), ecodriving is defined for thermal engine vehicles as following several driving advice such as:

- do not drive too fast,
- do not accelerate too quickly,
- shift gears sooner to keep engine speed lower,
- maintain steady speeds,
- anticipate traffic flow when accelerating and slowing down,
- keep the vehicle in good maintenance.

These advice are also subjective but the choice has been made to model this definition because it is the most common. This definition has the advantage to be rather simple to model. In this study, the attempt to model ecodriving will be focused on tuning driver parameters such as desired acceleration and speed, time headway,...

However, a complete definition should be more complex knowing that the driver practicing ecodriving has to consider other constraints than reduce fuel consumption.
1.1.2. Ecodriving definition proposal. Here, I propose a basis for the definition of ecodriving founded on a weighted balance of criteria. In figure 1, weights for 4 criteria are presented: safety, fuel economy, travel time and comfort. In fact, in real driving conditions, the way humans balance these criteria varies according to their driving styles (hypermiler, ecodriver, typical driver, sport driver). The curves on figure 1 define the evolution of weights with the driving styles and the pie charts below describe the weights repartition for four different driving styles:

- hypermiler: this category of drivers concentrates all its efforts on minimizing fuel consumption, trading off safety, comfort and travel time (Barkenbus, 2010). An example of a dangerous technique allowing to minimize fuel consumption consists in shutting down the engine in downhill.
- ecodriver: this category of drivers optimizes its commands to reduce fuel consumption without trading off safety. Examples of strategies to practice an ecodriving have been previously exposed.
- typical driver: represent the mean driver present on roads. This driver often preferred safety and comfort to fuel economy.
- sport driver: this category of drivers only considers its travel time without taking into account safety, comfort and fuel economy.

The way humans balance these constraints also varies according to their actions (navigation, control, stability) but this has not been represented in this study. In the next future, this definition will be employed to describe driving styles by building speed profiles. The idea is to compute four speed profiles, each of them dedicated to a criteria. Then, a global speed profile can be calculated using a weighted mean, depending on the driving style, of the four speed profiles previously computed.

1.2 Objective of this study
To provide the information about a whole population of drivers practicing ecodriving, the aim of this work is to study the impact of the penetration rate of ecodrivers, from 0% to 100%, on fuel consumption and traffic congestion. The impact on fuel consumption has been assessed through the quantity of fuel consumed and the impact on traffic congestion has been evaluated with the temporal mean speed of the vehicles.

The paper is structured as follows: section 2 presents the methodology of the study, while section 3 develops the modeling and section 4 exposes the results. Section 5 contains the discussions and sections 6, 7 and 8 are respectively the conclusion, the acknowledgments and the references.

2. METHODOLOGY

To analyze the impact of the proportion of ecodrivers on fuel consumption, different car-following and fuel consumption models have been used. The idea is to compute a speed profile for each vehicle that feeds into a simple fuel consumption model. For each numerical try, different combinations of the following parameters have been simulated: the number of vehicles constituting the population, the proportion of ecodrivers in this population, and the road type. The number of vehicles has been selected to represent free, intermediate and congested traffic state. These three situations have been tested on two types of road: inter urban and urban. For each study case, the proportion of ecodrivers has varied from zero to one hundred percent (by a step of 10) and each simulation has been replicated ten times. A total of 660 simulations were therefore performed. An advantage of testing a wide range of ecodriver proportions is that the effect of increasing the number of ecodrivers will be assessed. However, a drawback is that the current proportion of ecodrivers is unknown. To represent ecodrivers, a class of ecodriven vehicles has been defined with several parameters such as acceptable acceleration, time headway, maximum desired speed,... These parameters have been chosen on the basis of real world experiments (Saint-Pierre, 2010).

To compute the speed profile of a whole network, a car-following model has been employed on inter urban roads. The chosen model is the Intelligent Driver Model (IDM) (Treiber et al., 2000) as it takes into account the velocity differences between vehicles. This model has been slightly modified to represent the speed variation in curves where drivers are considered to travel at the $V_{85}$ operating speed (Louah et al., 2009). Other speed profiles could have been used (Glaser and Aguilera, 2003; Orfila et al., 2010) but the simplest has been selected in a first attempt. For urban situations, Aimsun simulation software has been used. This software, based on the Gipps car-following model (Gipps, 1981), can represent complex situations such as traffic lights, lane change, give way or roundabouts. In this study, the case of a district with
four traffic lights has been modeled. The computed speed and position of each vehicle through time have been recorded.

Using the precedent speed profiles, a fuel consumption model, founded on the mechanical energy consumed with a performance ratio function of the vehicle speed, computes the instantaneous and cumulated fuel consumption for inter urban roads. The performance ratio has been assessed using experiments on a specific car at various speeds and gear ratio (Wang et al., 2008). More complex models (Giakoumis and Lioutas, 2010) could not have been used but they will be tested in future works to take into account the engine speed. For urban roads, the Aimsun integrated fuel consumption model (Akçelik, 1983), has been used. For each study case, the level of traffic congestion has been assessed through the temporal mean speed of all vehicles.

3. Modeling

3.1 Modified car following model

In this study the IDM (Intelligent Driver Model) has been used for interurban situations. This model has the advantage to represent the selected drivers parameters and it also can be adapted to take into account the road curvature.

\[
\begin{align*}
\dot{x}_\alpha &= v_\alpha \\
\dot{v}_\alpha &= a \left( 1 - \left( \frac{v_\alpha}{v_0} \right)^\delta - \left( \frac{s_0 + v_\alpha T + \frac{v_\alpha \Delta v_\alpha}{2\sqrt{ab}}}{s_\alpha} \right)^2 \right),
\end{align*}
\]

where, \( x_\alpha \) (m) is the \( \alpha \) vehicle position, \( v_\alpha \) (m.s\(^{-1}\)) is the \( \alpha \) vehicle speed, \( a \) (m.s\(^{-2}\)) is the desired acceleration, \( v_0 \) (m.s\(^{-1}\)) is the desired velocity, \( s_0 \) (m) is the minimum distance between vehicles, \( T \) (s) is the desired time headway, \( \Delta v_\alpha = v_\alpha - v_{\alpha-1} \) (m.s\(^{-1}\)) is the approaching rate, \( b \) is the comfortable braking deceleration and \( s_\alpha = x_{\alpha-1} - x_\alpha - l_{\alpha-1} \) (m) is the net distance between vehicles with \( l_{\alpha-1} \) the length of the leader vehicle.

This model has been modified to represent the driver behavior in curves by adding a term to \( \dot{v}_\alpha \). The speed desired by drivers in curves can be assessed using the operating speed \( V_{85} \), defined as the 85th percentile of the speed distribution of free-moving cars (Vertet, 2006).
\[ v_\alpha' = a \left( 1 - \left( \frac{v_\alpha}{v_0} \right)^\delta - \left( s_0 + v_\alpha T + \frac{v_\alpha \Delta v_\alpha}{2ab} \right) \right) - \left( \frac{v_\alpha - V_{85}(x_\alpha + v_\alpha T_r)}{2T_r\sqrt{ab}} \right)^2, \]

where \( V_{85}(x_\alpha + v_\alpha T_r) \) is the operating speed at the distance \( x_\alpha + v_\alpha T_r \) and \( T_r \) is the headway time to the road events (curve,...). An example of speed profiles obtained for 50 vehicles is given in figure 2.

\[ \text{Figure 2: Vehicle speed versus distance} \]

### 3.2 Fuel consumption model

The theoretical consumed energy, \( E_{theo} \), in a time step \( dt \) can be evaluated by the following formulae:

\[ E_{theo} = \left( \frac{1}{2} \rho_{air} S C_x v_\alpha^2 + C_{rrr} m g + m p + m a_\alpha \right) v_\alpha dt, \]

where \( \rho_{air} \approx 1.2 \text{ kg.m}^{-3} \) is the density of the air, \( S (m^2) \) is the end face, \( C_x \) is the longitudinal drag coefficient, \( C_{rrr} = 0.015 \) is the coefficient of rolling resistance, \( m (kg) \) is the vehicle mass, \( g = 9.81 \text{ m.s}^{-2} \) is the standard gravity, \( p (\%) \) is the road grade, \( a_\alpha (m. \text{s}^{-2}) \) is the \( \alpha \) vehicle acceleration and \( dt \) (s) the time step.

Then, an efficiency ratio has been applied to take into account the energy lost in the combustion and transmission processes. This ratio comes from real experiments from Wang et al. (Wang et al., 2008) where the fuel consumption has been measured at different speeds.
\[ \eta = \frac{E_{\text{theo}}}{E_{\text{meas}}} = 10^5 \times \frac{1}{f(v_a)} \frac{\rho_{\text{air}} S C_m v_a^2 + C_{rr} m g + m p}{e_{\text{carb}} \rho_{\text{carb}}} \]

where \( E_{\text{theo}} \) is the theoretical energy consumed if the vehicle was the one tested by Wang et al., \( E_{\text{meas}} \) the measured energy consumed by the test vehicle, \( f(v_a) = 0.05v_a^2 - 1.8v_a + 21 \) is a function fitted on the works of Wang et al. giving the fuel consumption in liters per hundred kilometers versus the vehicle speed, \( e_{\text{carb}} = 42.5 \times 10^6 \text{ J.kg}^{-1} \) is the energy density of fuel and \( \rho_{\text{carb}} = 0.84 \text{ kg.L}^{-1} \) is the fuel density. The computed efficiency ratio can be seen on figure 3.

![Figure 3: Efficiency ratio](image-url)

4. RESULTS

Table 1: Drivers parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal driver</th>
<th>Ecodriver</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Sd</td>
</tr>
<tr>
<td>Desired speed ( v_0 )</td>
<td>30</td>
<td>5</td>
</tr>
<tr>
<td>Headway time ( T )</td>
<td>1.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Road headway time ( T_r )</td>
<td>3</td>
<td>0.9</td>
</tr>
<tr>
<td>Desired acceleration ( \alpha )</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>Braking deceleration ( b )</td>
<td>3</td>
<td>0.9</td>
</tr>
<tr>
<td>Minimum distance ( s_0 )</td>
<td>2</td>
<td>0.6</td>
</tr>
</tbody>
</table>
Simulations have been run with different driver parameters for normal drivers and ecodrivers. For each virtual driver, parameters have been randomly generated using values from table 1 with a Gaussian distribution. To each parameter is associated a mean value, a standard deviation (Sd), a minimum and a maximum value.

4.1 Interurban cases

To analyze the impact of ecodriving in interurban situations, a theoretical road has been designed. It consists in a straight line and a curve every 1500 meters introduced and concluded with a clothoid. On this road, three traffic conditions have been applied from a free traffic to a congested one. These three situations have been arbitrarily determined on the assumption that, in a free traffic, vehicles reach their maximum speed and in a congested traffic, vehicles travel very slowly. For interurban cases, the situation where vehicles are stopped has been avoided because the fuel consumption model does not take this case into account. The third situation is an intermediate state between the free one and the congested one. The fuel consumption, given in liters per 100 km, is the mean consumption of all vehicles for the ten numerical tries. To analyze the impact on traffic, the mean speed has been selected.

4.1.1. Free traffic state. This situation has been reached with 10 vehicles traveling on the virtual road. In figure 4, fuel consumption is steadily decreasing from 9.05 to 8.35 L/100 km, that is 7.7%. The solid line represents the mean and the dotted lines represent the standard deviation on the ten replications. Mean speed is also decreasing but in a quadratic way from 60.5 to 55 km/h, that is 9.1%. This is equivalent to 2 minutes lost on a 20 minutes travel. For this situation, it could be said that increasing the proportion of ecodrivers, as defined previously, is a good way to reduce fuel consumption without loosing too much time.

4.1.2. Intermediate traffic state. This situation has been reached with 50 vehicles travelling on the virtual road. In figure 5, it can be seen that fuel consumption drops from 8.2 to 7.5 L/100 km.
km, that is 8.5%, between 0 and 20% of ecodrivers. Then, it increases to reach 8.4 L/100 km for 100% of ecodrivers, that is 2.4% more than without any ecodriver. This can be explained by the mean speed, which is sharply decreasing from 47 km/h to 29 km/h, a difference of 38%. Indeed, when speed drops sharply, the efficiency ratio of the engine also drops sharply from 0.13 to 0.06, doubling the lost energy. This effect should not appear for electric vehicles. This sudden drop in speed could be explained by the traffic instability occurring when the traffic goes from a free state to a congested one. In this particular situation, increasing the percentage of ecodrivers above 20% is not recommended and trying to reach this optimum of 20% of ecodrivers warrant a decrease in speed of only 7.5%.

4.1.3. Congested traffic state. This situation has been reached with 100 vehicles travelling on the virtual road. In figure 6, the fuel consumption decreases rapidly from 13.4 to 10.9 L/100 km, that is 18.6%, between 0 and 30% of ecodrivers. Then, it decreases steadily to reach 8.6 L/100 km. This traffic state is the one where eco-driving has the strongest effect as the total gain is about 35.8% between 0 and 100%. However, in the same time, the speed has decreased of 55.1%, doubling the travel time. A good compromise could be the proportion of 30% of ecodrivers where fuel consumption has been reduced of 18.6% while the speed has decreased of 23.7%.
4.2 Urban cases

Urban cases have been treated with Aimsun simulation software (from TSS company) to represent a complex situation such as a district with four traffic lights. In figure 7, the three tested conditions are represented. Light gray vehicles are ecodriven and dark ones are normally driven. The fuel consumption is given in liters and is defined as the quantity consumed by all vehicles that have finished their trip during the simulation process. The speed is the mean speed for all vehicles that have left the network.

Figure 7: traffic states: left, free; center, intermediate; right, congested

4.2.1. Free traffic state. This situation has been reached with a traffic flow of 100 vehicles per hour for each of the 8 entrances of the network. In figure 8, the fuel consumption decreases slowly from 3.55 to 3.2 liters, that is 9.8% and the speed decreases from 26.9 to 21.6 km/h, that is 19.7%.

Figure 8: left, fuel consumption for a free traffic state; right, speed

4.2.2. Intermediate traffic state. This situation has been reached with a traffic flow of 400 vehicles per hour for each of the 8 entrances of the network. In figure 9, the fuel consumption seems to increase from 37 to 38.5 liters then decreases to 34.2 liters for 100% of ecodrivers.
The fuel economy between 0 and 100% of ecodrivers is about 7.6%. The mean speed decreases from 6.5 to 4.4 km/h, that is 32.3%. Finally, in this situation, the proportion of 40% should be avoided for a fuel economy purpose and the proportion of 80% and above should be avoided for a travel time purpose.

![Figure 9: left, fuel consumption for an intermediate traffic state; right, speed](image)

4.2.3. Congested traffic state. This situation has been reached with a traffic flow of 800 vehicles per hour for each of the 8 entrances of the network. In figure 10, fuel consumption decreases slowly from 0 to 20% of ecodrivers then drops to 37 liters for 100% of ecodrivers what represents 19.6% of economy. The mean speed is very lightly increasing around 4 km/h. In this situation ecodriving could be generalized to all drivers and this situation is the one where ecodriving is the more efficient.

![Figure 10: left, fuel consumption for a congested traffic state; right, speed](image)

5. DISCUSSIONS

In this study, two road infrastructures with three traffic states have been tested. Results for inter urban roads show that:
• increasing the proportion of ecodrivers is efficient on a free traffic state as the fuel consumption decreases significantly while the mean speed varies in the same order.
• for the intermediate traffic state, an optimum proportion of 20% of ecodrivers has been found. Increasing the percentage of ecodrivers above this threshold degrades the situation because the speed suddenly drops.
• for the congested traffic state, a compromise has been highlighted for 30% of ecodrivers. At this level, the fuel consumption has been reduced and speed has been maintained at a suitable value.

Results for urban roads show that:
• for a free traffic state, speed and fuel consumption are linearly decreasing with the proportion of ecodrivers. In these conditions, the optimal proportion of ecodrivers depends on the politicians and their choice to encourage ecology or traffic fluidity.
• For the intermediate traffic state, the fuel consumption increases from 0 to 40% of ecodrivers and the optimum should be about 60% because the speed is very low above. Proportions below 50% and above 80% should be avoided for a traffic fluidity purpose.
• For the congested traffic state, ecodriving is very efficient as the fuel consumption strongly decreases while the speed is almost steady.

In this work three main results have to be highlighted:
• Firstly, ecodriving is, in most of simulated cases, an efficient way to reduce fuel consumption from 7.6 to 35.8%.
• Secondly, in some situations such as the intermediate traffic state on an inter urban road, the fuel consumption has increased with the proportion of ecodrivers. This unexpected result, mainly explained by the engine efficiency ratio and by the traffic instability, proved that in particular cases, generalizing ecodriving could be counter-productive.
• Thirdly, the optimal proportion of ecodrivers is different from a situation to another. In some situations, drivers cannot know if they have to practice ecodriving or not because they do not have the information of the optimal proportion of ecodrivers nor the current proportion of ecodrivers.

On the basis of these results, three solutions can be proposed:
• a short term one consisting in promoting ecodriving only in clearly identified situations such as totally free and congested traffic. Ecodriving learning processes should include this consideration.
• a middle term one consisting in informing drivers on the driving style they have to perform, for example by the way of variable-message signs.
• a long term one consisting in optimizing driver assistance systems such as ACC (Adaptive Cruise Control) according to the driving conditions. Information concerning
the traffic state and the road infrastructure could be exchanged by vehicles and infrastructure in order to compute the optimal parameters for all vehicles in the network. Works in this direction, for an isolated vehicle in realistic traffic conditions (Kamal et al. 2010) and for a whole network (Barth and Boriboonsomsin, 2009), have already started.

It should be kept in mind that, in this study, some hypothesis have been made and must be considered in the analysis of results:

- Efficiency of ecodriving highly depends on the relative importance that drivers set on fuel economy and time spent. Previous analysis of results are valuable on the hypothesis of an equivalent importance of these two criteria.
- Although it is one of the main criteria of drivers, the effect of the proportion of ecodrivers on safety has not been assessed in this work.
- The impact of ecodriving on traffic congestion has been studied through the mean temporal speed of ten numerical tries. This could be completed by analyzing the mean speed of the 10% of vehicles that are the most delayed in comparison with a free traffic state.
- In the results, fuel economy is measured in comparison with a situation without ecodrivers. However, an unknown proportion of ecodrivers is already in the population and it is unreasonable to estimate that 100% driver will ecodrive tomorrow.
- Although it is an important parameter of ecodriving, the engine speed has not been taken into account in the fuel consumption models. This choice, that has been made to simplify the model, could have an effect, especially for congested traffic states which imply numerous gear changes. Furthermore, experiments used to specify the driver parameters were not dedicated to the car-following models used in this work.
- The only tested vehicles are thermal engines vehicles. It could be interesting to test other vehicles such as electric vehicles with an efficiency ratio less dependent on the speed and a possibility to regenerate energy while braking. Also, trucks and buses, which can represent an important part of fuel consumption, have not been modeled.

6. CONCLUSIONS

This paper has studied the impact of the penetration rate on fuel consumption and traffic congestion by running simulations on different situations. According to these simulations, the optimal proportion of ecodrivers is radically different from a situation to another. For a congested traffic, ecodriving, as defined in this study, is almost beneficial while it could be cons-productive with an intermediate traffic state on inter urban roads. As this cons-productive effect is rather limited (only 2.5% of increase in fuel consumption) for a traffic state, a priori, less frequent than others, it should still be advised to encourage ecodriving.
Theoretically, this study will have an impact on the way ecodriving is considered in our research. In practice, these results can be implemented in three ways:

- promoting ecodriving only in situations such as totally free and congested traffic. Ecodriving learning processes should also include this consideration.
- informing drivers on the driving style they have to perform, for example by the way of variable-message signs.
- designing driver assistance systems that could communicate with the infrastructure to evaluate the traffic state and so, automatically adapt the vehicle behavior. This can be performed with Adaptive Cruise Control (ACC) systems automatically parameterized by the network according to the driving conditions (road, traffic state,...).

Future works will be concentrated on validating these simulations experimentally and on improving the fuel consumption and car-following models to take into account drivers safety. The optimization process allowing to find the best driver parameters according to the driving conditions will also be studied.

7. ACKNOWLEDGMENTS

The author would like to thank the ANR-09-VTT-01 for funding this research in the ABV project.

8. REFERENCES

Glaser, S. and V. Aguiléra (2003). Vehicle-Infrastructure-Driver Speed Profile: Towards the Next Generation of Curve Warning Systems, 10th World Congress and
Exibition on Intelligent Transport Systems and Services (ITS).


MONITORING OF LANDSLIDES ON THE PAN-EUROPEAN CORRIDOR X FOR THE PURPOSE OF ENVIRONMENTAL PROTECTION
ABSTRACT

On the section of the highway E-75 passing through Serbia, which lies within the Pan-European corridor X, deformations occurred after finishing the construction of the road. During the time, rate of deformations has become intensive which jeopardized traffic operations. The most intensive movements occurred on five locations - Beska, Begaljicko brdo, Kolari, Bracin and Razanj, where landslides were formed. The monitoring system was established for the purpose of solving causes of their activations. It consisted of inclinometer and piezometer constructions and geodetic marks, which were monitored for a long period of time. Besides the results of monitoring, this paper presents numerical model established for the purpose to indicate the potential causes of the Kolari landslide deformation history. Certain conclusions are derived based on measured data and previously made assumptions. All analyses are performed by using limit equilibrium (LEM) and finite element methods (FEM).

**Keywords:** Corridor X, landslide, monitoring, deformations, limit equilibrium method (LEM), finite element method (FEM), environment, Plaxis

**Research domains:**
Transport sustainability and environment
1 INTRODUCTION

Highway E-75 is integral part of Pan-European corridor X and it is the main route in the Republic of Serbia. It connects its Northern part (border with the Republic of Hungary) with its Southern part (border with the Republic of Macedonia). The Northern part of the route passes through Autonomous Province of Vojvodina. On the section from Hungarian border to the city of Novi Sad construction of new lane is currently underway. On the section from the city of Novi Sad to the entrance to Belgrade, highway already exists in its full width (with physically separated lanes). Most of the Autonomous Province is a lowland, thus the alignment to be set is not a challenging task. Far bigger problems can occur due to setting the alignment through hilly and mountainous regions of Central and Southern Serbia. In this part, the highway should be constructed at the location of Belgrade bypass, and on the section from the city of Leskovac to the Macedonian border, Figure 1. With the construction of existing highway some land-mass movements occurred in several locations, which jeopardized traffic operations. The large-scale movements occurred on five locations: Beska, Begaljicko brdo, Kolari, Bracin and Razanj. Due to the importance of the road route the landslide monitoring was established on four of those locations (Beska, Begaljicko brdo, Kolari and Razanj).

![Figure 1: Geographical position of major landslides on Corridor X](image)

1.1 Problem definition

Problems concerning traffic operations can often be caused by land-mass movements. In order to solve those problems adequately, produced movements should be classified according to
their velocity. Varnes (1958) proposed classification system that divides landslides into 7 categories ranging from extremely slow creep (max 6 cm/year) to extremely fast moving landslides (more than 3 m/sec.). [17]

Landslide that experienced rapid movements occurred near village of Bracin (main scarp was 4.0 m high), Figure 2a. Certain repair measures were undertaken, but these improvements didn’t produce the effects as expected. Because of the amount of moved mass this landslide was repaired by passing the bridge over it, Figure 2b. Necessity for solving above-mentioned problems in short periods of time is occasionally producing inadequate repair measures, thus leading towards new repair designs with more expensive solutions.

On the other hand, landslides involving small rate of movements (Beska, Begaljicko brdo, Kolari and Razanj) can jeopardize traffic operations only after longer periods.

Occurrence of slow rate movements can be triggered by different factors. Some movements are observed to develop under constant (or nearly constant) water level and are studied by laws of viscoplasticity. On the other side, large number of movements is caused by different hydrological factors. They are usually leading towards development of higher pore water pressures in landslide body. Latter is recognized as one of the most common factors influencing the movements. Hydrological factors can produce positive pore pressures thus lowering effective stresses and shearing resistance. They can also decrease negative pore pressures in unsaturated soils thus neglecting effects of suction.

Due to the complex nature of these movements, adequate repair measures are usually difficult to determine. In order to eliminate uncertainties prior to making conclusions, the monitoring should be established. In this way, interpretation of results is more objective.

Mentioned landslides are being observed since 2002, except for landslide Begaljicko brdo, which is being monitored since 2005 and landslide Beska which is being monitored since 2007. Procedures concerning long-term observations can differ significantly by type, size and methodological approach. However, their objectives and tasks are always the same: to examine sliding mechanism and dynamics that are crucial factors in determining the most applicable repair measures.

Next, this paper will present the monitoring methodology conducted on Kolari landslide. Further on, numerical model, based on finite element method (FEM), will be introduced to try to prove or disapprove influence of certain hydrological aspects concerning movements of landslide. Numerical model was also used to obtain global factor of safety (FoS) and to compare it with FoS from monitoring design.
2 KOLARI LANDSLIDE

Landslide is jeopardizing traffic operations near to Kolari village on Highway E-75, i.e. section Belgrade - Nis, from km: 636+125.89 to km: 636+453.31, Figure 3. Since the end of September 2002, monitoring system has been established by installation of geodetic marks, inclinometer and piezometer constructions. Monitoring was being performed from October 2002 till April 2009.

At the location under study the road alignment was set in the cut with maximum height of 8.0 m. Initially, right lane with cut slopes, V:H=1:2, and retaining wall were constructed. With the widening of the road to full-fledged highway profile, cut slopes were slightly sloped to V:H=1:3. At the back of retaining wall, the drainage system was installed and connected to revision shafts and drains being linked for the discharge into the Ralja river. Due to deformations that occured at the beginning of the cut, retaining wall was extended in direction towards Belgrade for 6.0 m.
In the coming years, deformations appeared on several occasions and different repair measures had to be performed (repair and strengthening of drainage system and replacement of pavement structure). These improvements didn’t produce the effects as expected. In order to solve the emerging issues the monitoring system was established. Since 2002, 19 geodetic marks have been installed, 23 inclinometer and piezometer boreholes. Today, only 7 inclinometer and 7 piezometer constructions and 9 geodetic marks are in function. [16]

2.1 Geotechnical investigations and monitoring

Aerial photograph shows the position of performed field works since 2002, Figure 4. In order to carry out the stability calculations and subsequently repair measures, laboratory testing on soil samples had to be performed and for that purpose 110 soil samples were tested.

Figure 4: Position of geotechnical field works with contours of the landslide

2.1.1 Geological characteristics

Field investigations have helped to establish subsequent geological profile:

- Colluvium (moved mass) that has varying thickness and heterogeneous composition. It consists of loess like sediments (ML) - Layer #1, clayey silt (MH) - Layer #2, clay with gravel - Layer #3 and clayey sand (SM-SC) - Layer #4. Pavement structure - Layer #10 and road foundation - Layer #9 are encompassed by ground movements also. Colluvium has maximum thickness, of 17.0 m, at the top of cut slope, and varies from 6.0 to 8.0 m in the zone of highway.

- Fill material - Layer #5 is formed from the excavated material.

- Bedrock layer - Layer #6 consists of neogene sediments. They are vertically changing from sands to silts.
Alluvium deposits consist of silty clay (CH) - Layer #7 and sand (SW/SC) - Layer #8.

Borehole core samples are shown on Figure 5.

Figure 5: Borehole core samples

Landslide description. Kolari landslide is old deep-seated landslide formed on the right side of Ralja river valley in landslides-prone area. One of the main causes for forming the landslide is river erosion. Main triggering factor for its reactivation is recognized as seasonal fluctuation in rainfall.

Main scarp is located at about 80.0 m distance from the retaining wall, on the slope above the alignment (on elevation of 115 mOAD). Toe of the landslide is situated in the zone of the left lane ditch, close to the alluvial sediments of Ralja river (on elevation of 100 mOAD). According to the measurements, largest movements were found on the slope just above the retaining wall. Large movements were also measured at the location of the wall, and in the zone of the right lane. Moderate deformations in comparison to the previous ones were recorded within the left lane zone.

Inclinometer measurements detected the presence of two shear planes. Position of these zones is in accordance with engineering geological borehole core mapping. Main shear plane (denoted in monitoring design as “deep shear zone”) is formed at the interface of Layer #4 and Layer #6. The second shear plane (denoted in monitoring design as “shallow shear zone”), being mainly affected by human activity on the ground surface, is found to be at depths ranging from 2.0 to 3.0 m. Position and shape of shear planes are shown on cross-section as per Figure 6.
Figure 6: a) Engineering geologic cross-section
b) Phase construction
c) Model for time dependent numerical analysis
2.1.2 Monitoring results

Deformation measurement results. Induced displacements in inclinometers, during monitoring, amount to approximately few centimeters (4.0 to 5.0 cm). Magnitude of displacements points out the fact that these movements represent extremely slow creep.

In regions where rainfall-triggered landslides predominate, stick-slip movements of relatively small magnitude in some areas are not uncommon. However slow the movement, one is often dealing with an existing landslide and that fact will have a great bearing on the scope and goals of analysis. Landslide movements of the order of millimeters per year, (in the slow to extremely slow range), are very difficult to detect at first and require local knowledge and careful investigation such as a study of aerial photographs and search for subtle local effects such as minor cracking. Monitoring with subsurface instrumentation (for example, use of inclinometers) is often necessary under such circumstances (Chowdhury 2010, pp. 115-116).

Inclinometers showed that all measuring points move towards the river (axis-A) and towards the city of Nis (axis-B). Figure 7 shows plots of incremental and cumulative displacements, and vector of total displacements along the vertical profile of the inclinometer construction IB-4. Construction was installed in 2002, and its spatial position is shown on Figure 2. Total of 26 series of measurements were performed, but in this case only the measurements at the end of each year are shown.

On the plot of total (cumulative) displacements two shear planes could be distinguished at depths of 2.0 and 6.90 m. Overall displacements at the depth of 6.90 m (direction of axis-A), are somewhat more than 2.0 cm. Shallow zone has the displacements of about 3.0 cm. Whereas those in direction of Axis-B are much smaller (about 0.3 cm).

Ground surface movements reach their maximum at the slope of cut (18.0-35.0 cm). Geodetic marks in the central reserve show slightly smaller movements (3.0-4.0 cm).
Groundwater level measurements. In order to perceive the conditions and regime of groundwater within the terrain, a monitoring network of piezometers (open standpipes) was established. Long-term measurements showed that there are two aquifers which are not hydraulically connected. Piezometer constructions are installed by means of batteries of piezometers inserted at different depths. Measured water pressures in the body of landslide indicated the existence of unconfined aquifer. These measurements indicated onto the seasonal variations that are within the range of 4.50 m, Figure 8.

Deeper piezometers measure water levels in lower aquifer, yet the latter being somewhat confined. Results pointed out that the lower groundwater level oscillations are within the range of 1 m. At the toe of the landslide this aquifer is becoming unconfined and hydraulically connected with the waters of the Ralja River. [16]
Piezometer constructions installed in the zone of pavement structure indicated the minimum or no oscillations whatsoever (variations have been between elevations of 98.54 and 99.03 mOAD). This means that the road foundation is constantly water-saturated, thus heaving of ground in the zone of pavement structure being the cause for deformation occurrence could be eliminated. [16]

*Global stability analysis.* Global stability analysis in the monitoring design was performed according to Janbu’s simplified method. Back-analysis calculation of the shear strength parameters was performed by assuming equally distributed shear strength parameters along weakened zone. By taking that the cohesion is close to zero, determined value of angle of internal friction equals $\varphi = 7.5^\circ$. This is within the range of measured values of residual shear strength parameters ($\varphi$ is ranging from 6-9°, and cohesion equals zero), as stated in laboratory report. Analyses were also performed for conditions before road construction. For residual shear strength parameters, obtained factor of safety was $FOS = 1.3$, which could classify the terrain as conditionally stable. This has led to conclusion that the construction of the highway undermined the stability of the slope. [16]

## 3 DISCUSSION

### 3.1 Numerical model

Geotechnical model of the terrain was implemented in *Plaxis* FEM code. So far, stability analysis had been performed on the basis of steady-state water conditions (maximum or minimum level) without considering time effects. Due to the fact that the landslide is rainfall triggered, key-role on performing adequate analysis has time factor. Purpose of this analysis is
to determine the influence of seasonal variations of water level on magnitude of deformations that were measured in the field. In order to account for all relevant factors, semi-coupled hydro-mechanical approach was applied. This method is based on Terzaghi’s principle of effective stresses, which makes the distinction between effective stresses and pore pressures.

Considered model consists of 1643 finite elements (with 14135 nodes) with average element size of 1.80 m. Accuracy of FEM is controlled by generated number of finite elements (more elements - higher accuracy). In this case, fine mesh was utilized, thus calculation then becomes more time consuming, i.e. less robust. Generated mesh is unstructured with 15-noded triangular elements (additional PlaxFlow module that was used for transient groundwater flow calculations uses 3-noded elements). In order to achieve compatibility between Deformation and Flow module PlaxFlow simply transform 15-noded element on sixteen 3-noded elements with one Gauss stress point per element. After flow calculations these 3-noded elements are transformed to 15-node elements by interpolation of Gauss stress points to original twelve as for 15-node used in Deformation module. While calculating transient groundwater flow stress state and deformations of soil skeleton are not considered. Calculations are performed with hydraulic parameters only, determined on the basis of Van-Genuchten hydraulic model. This model allows for calculation of suction dependent tensile pore pressures (these tensile pressures are neglected when calculating deformations, which is safer approach).

Field investigations had shown that it is applicable to present the model in two dimensions considering plane-strain conditions. Influence of lateral sides (which could be taken into consideration in 3D model) is not taken into account which is assumed to be appropriate due to the shape of landslide.

Two soil models are introduced in the analysis. Linear-elastic model is simulating behaviour of retaining wall and pavement structure. Hardening soil model is used to simulate soil behaviour, and is based on non-linear stress-strain relationship as observed for soils. Furthermore, it is based on the theory of plasticity rather than theory of elasticity as it is a case for the firstly mentioned model.

All calculations have been performed in drained conditions which is reasonable because there does not exist any rapid change in loading conditions, and landslide material is creeping extremely slowly, thus it is impossible for undrained conditions to prevail.

For conditions before road construction and for present conditions, safety analyses for obtaining the most critical failure mechanism were performed. Those analyses are based on strength reduction approach, firstly introduced by Zienkiewicz et al. (1975). [1], [18]

The calculation procedure was implemented in several phases, as shown in Table 1.
Changes of slope geometry are shown on Figure 6. The first phase encompasses the excavation of material for the right lane with cut slopes of V:H=1:2. Further, fill material is placed followed by gentle sloping, construction of pavement structure and retaining wall. After that, road is widened to its full width which represents current conditions. This construction process in phases allows us to determine global FoS (phases 6, 7, 8, 9) and history of induced deformations for every phase with shear zone taken into account. In order to try to back-calculate deformation history of landslide it is necessary to reset accumulated displacements from previous phases to zero in phase no. 5. In this way all relevant factors could be considered from the moment of installing the inclinometers. This phase is considering the transient groundwater flow.

<table>
<thead>
<tr>
<th>phase no.</th>
<th>phase name</th>
<th>type of calc.</th>
<th>loading cond.</th>
<th>water cond.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>initial conditions</td>
<td>total multiplier</td>
<td>self weight</td>
<td>steady state</td>
</tr>
<tr>
<td>1</td>
<td>activ. of sliding plane</td>
<td>staged constr. drained</td>
<td>no change</td>
<td>steady state</td>
</tr>
<tr>
<td>2</td>
<td>excavation 1</td>
<td>staged constr. drained</td>
<td>removal of the material</td>
<td>steady state</td>
</tr>
<tr>
<td>3</td>
<td>excavation 2 and loads</td>
<td>staged constr. drained</td>
<td>remov. and traffic load</td>
<td>steady state</td>
</tr>
<tr>
<td>4</td>
<td>excavation 3 and loads</td>
<td>staged constr. drained</td>
<td>remov. and traffic load</td>
<td>steady state</td>
</tr>
<tr>
<td>5</td>
<td>variation in WT</td>
<td>staged constr. drained</td>
<td>WT change in time</td>
<td>transient</td>
</tr>
<tr>
<td>6,7,8,9</td>
<td>strength reduction</td>
<td>incremental multiplier</td>
<td>phi-c reduction</td>
<td>steady state</td>
</tr>
</tbody>
</table>

3.2 Numerical analysis

3.2.1 Analysis of hydrological factors

In order to achieve pore water pressure distribution as measured and shown on Figure 8, variations in hydraulic head have been introduced. It should be mentioned that this groundwater level is hydrostatic and by introducing transient groundwater flow, hydrostatic state does not exist anymore. Differences are dependant on the magnitude of filtration forces (they are usually very small and can be neglected).

As it can be seen from Figure 8, oscillations of water level in piezometer constructions are not consistent, i.e. while some open standpipes show maximum values other show minimum in same period. Laboratory tests on soil samples showed very low hydraulic conductivity \( k_{sat} \) of soil layers in the body of landslide. This fact wasn’t considered by the time of installation of piezometers and has led to inappropriate measurement of water level oscillations. In soils with low permeability time that is needed for the water level to reach steady state in piezometer can be relatively long thus, open standpipe can be late with the reaction due to the water level change. It is uncommon to expect that large oscillations of water level can be achieved in clayey type soils with low permeability.
In order to prove abovementioned statement oscillation of hydraulic head was introduced in numerical model. As referent distribution of pore water pressures piezometer construction Pb-4 was considered, due to the highest oscillations in water level.

After transient calculations oscillations of water level according to Pb-4 could not be achieved with numerical model. Obtained variations of pore pressures were considerably smaller (max 5-10 kPa) than expected (45 kPa). Measured values were probably influenced by certain side effects (infiltration of water from the ground surface, human activities, etc.). They are characteristic just for piezometers and cannot be adopted as global. Realistic variations of water level could be reproduced by using infiltration option in PlaxFlow code. With this option daily amount of precipitation quantity can be incorporated for the period of monitoring (day by day for seven years). This is very extensive work and will not be examined in this paper (but will be examined in the future).

### 3.2.2 Safety analysis

Numerical model was also used to perform safety analyses. Safety analyses have been performed after every construction phase to investigate the influence of change of geometry on FoS. Some advantages of numerical model have been used for more reliable calculation of values of FoS.

Inspection of borehole core samples showed that the thickness of the weakened zone was approximately 10.0-20.0 cm. Such a thin layer (compared to scale of the landslide) is difficult to simulate by means of FEM, so interface elements can be utilized instead. Thickness of these elements is calculated as the virtual thickness factor times average element size (virtual thickness factor was chosen as 0.1 and by multiplying it with average element size gives value of 18.0 cm which is as observed for shear plane thickness). Details concerning characteristics of these elements could be found in [2].

In this particular case shear plane is passing through several different soil layers, thus generalization of shear strength profile as made in monitoring design does not make any sense. Based on investigations and by taking into account time of landslide appearance and its mechanism, it is assumed that along shear plane one may adopt residual shear strength parameters. Along weakened zone adequate residual shear parameters were assigned with reference to soil layers.

Main question in safety analysis was: Which water level to adopt as referent value? Due to the fact that landslide is rainfall triggered realistic variations of water level could not be reproduced exactly (as mentioned before). In order to compare FoS between two analyses same referent water level was used as in monitoring design.
Safety analysis showed that the global factor of safety, for present conditions, equals FoS = 1.14, Figure 9. This means that slow ground movements are performed with this factor.

It should be noted that the limit equilibrium method does not provide forecasts of magnitude of deformations concerning different factors of safety, except warning that in case of FoS ≈ 1 displacements are going to be large. For FoS > 1 magnitude of displacements decreases with the increase of safety factor, but the actual magnitude of displacements depends on the stress-strain characteristics of soils (Maksimovic 2005, pp. 442). [10]

More realistic value of FoS could be obtained by introducing oscillations of water level. FoS would than vary between maximum and minimum value that is related to the oscillation magnitude. As mentioned earlier water level oscillations are not expected to be great, thus variations of value of FoS should be close to 1.14.

Factor of safety that was obtained in the monitoring design, for conditions before road construction, showed value of FoS = 1.3. Analysis performed based on strength reduction approach showed larger value of FoS = 1.86. Differences are due to the fact that numerical model was established by introducing thin interface zone with residual shear strength parameters (as it was observed on the field), and limit equilibrium approach was performed with Janbu’s simplified method, which doesn’t give reliable results for deep landslides, that consist of heterogeneous material. [10]

Further, in monitoring design residual strength parameters were assumed to be equal all over the sliding surface which is not appropriate.

From previous analyses another important question arises: Is value of $\varphi = 7.5^\circ$ as obtained in monitoring design correct if knowing the fact that method used for analysis is not appropriate one? Answer to that could be found if performing back-analysis with some method that satisfies all elements of statical equilibrium. Value of $\varphi$ should be smaller, i.e. closer to lower bound of $6^\circ$ as measured in laboratory (this statement will be examined in the future).

For the cases involving phase construction, FOS had values between 1.50 and 1.56 until the beginning of excavation for the construction of left lane. Removed material, triggered slow movements as one can observe today.
4 CONCLUDING REMARKS

This paper shows some aspects of monitoring landslides. Results stated in monitoring design were used as a beginning basis. Those results have been tested by implementing numerical model.

Main unknown represents transient water level. Further investigations are needed to fully prove in-situ conditions. Correct solution would be to introduce precipitation effects by collecting data from local hydrometeorological service. In this way, realistic oscillations of water level with time and corresponding pore pressures could be obtained.

These hydrological effects have influence on FoS. Safety analysis showed that the global factor of safety, for present conditions, equals FoS = 1.14. More realistic value of FoS could be obtained by introducing oscillations of water level. FoS would than vary with every change of pore water pressures.

Safety analysis with numerical model confirmed the fact that Janbu’s simplified approach is not appropriate method for analysis of deep heterogeneous landslides. This fact should also be proved by utilizing LEM that satisfies all elements of statical equilibrium.

Given the large number of laboratory tests on soil samples, stress-strain characteristics can be considered to be determined reliably, while hydraulic parameters are determined mainly based on empirical data and literature. In the future, it would be desirable to confirm experimentally all parameters, especially hydraulic, because the selection of parameters greatly affects final deformation characteristics of the model.
One of the drawbacks of numerical model represents the fact that in the calculation (taking into account the flow of water) one can not simulate the behavior of deeper confined aquifer.

Influence of the river Ralja was not considered because its riverbed has already been regulated.

The advantage of established model, compared to conventional, is that on the basis of these results one can provide a reliable proposal for the most optimal repair measures, which can ensure traffic safety.

5 References


Juned Akhtar*
* Corresponding author (jak@toi.no). Telephone: 0047 22573823. Fax: 0047 22609200
Institute of Transport Economics (TØI), Gaustadalleen 21, NO-0349 Oslo, Norway

Oil spill risk analysis of routeing heavy ship traffic in Norwegian waters

May 2011

Abstract

Norwegian authorities have for a long time been concerned about the risk of oil spills outside the Norwegian coast. One of the key measures adopted have been to reduce the risk of ship accidents by imposing sailing routes for heavy ship traffic (over 5,000 GT) with high environmental risk potential farther away from part of the coast. This article is based upon two reports which conducted risk assessments of imposing such sailing routes outside the entire Norwegian coast. These routes were proposed by an expert group consisting of relevant stakeholders.

Data of traffic pattern and number of sailing were collected for the year 2008 using the universal Automatic Identification System – AIS. The proposed route was compared with today’s traffic pattern in regards to the accident frequencies and the expected oil spills per year. An accident and oil spilling simulation program called MARCS was used to simulate these results. After conducting a traffic forecast for the year 2025, the simulation was again run and the results compared with the year 2008. In total, the proposed routing is expected to reduce oil spills by 590 tonnes per year in 2008 and by 3670 tonnes in 2025. The main reason for this substantial reduction is that the number of groundings is reduced because of the distance from the shore being increased. The reduction was particularly strong for tankers.

Keywords: Marine traffic, environment, simulation, accident probabilities, oil spill
Research domain: Transport sustainability and environment
1. Introduction

Commercial shipping has important impacts on the wider environment, due to the ordinary release of exhaust gases etc., but in particular due to the risk of accidents with the unintended release of toxic chemicals and oil spills. Unfortunately shipping accidents may have very severe negative impacts, particularly to coastal regions, due to the potential release of very large quantities of hazardous or eco-toxic cargo materials such as crude oil.

Since 2002 the oil transport in the Barents Sea from Russia has increased significantly. In 2002, 4 million tonnes of oil were shipped westward through the Barents Sea. In 2008 the number had increased to 10.8 million tonnes. Forecasts for 2025 estimate a 60 per cent increase in the oil transport and a tenfold increase in the gas tanker transport (Hovi & Madslien 2008; Frantzen B. and A. Bambulyak (2008).

The Norwegian Sea is rich on natural resources, thus being of great economical interest. Along with the fishery resources it also has huge resources of gas and oil, resulting in high transport activity. Recently launched projects such as the gas production from the world’s largest offshore gas reserve, Shtockman, and the production of LNG from Snøhvit and oil from Goliat will add to today’s already high maritime activity along the coast of Norway and thus increase the risk of environmental damages. The marine flora and fauna are vulnerable; the Norwegian coast line carries approximately 19.5 million individual sea birds and 4.5 million breeding pairs (Loeng, H. and K. Drinkwater, 2007)

All oil transport imposes a risk of acute oil pollution. Consequently an increase in the maritime and offshore petroleum activity will increase the risk. Experience shows that only 10-15% of oil spills in the Arctic Sea can be removed by the current level of preparedness (Frantzen B. & A. Bambulyak (2008). The rest will have to be left to the natural evaporation or breakdown over time. The major contributor to this risk is tankers transporting oil from Russia along the Norwegian coast (Kystverket, 2006).

Since 2000 Norway has experienced several adverse events that could have resulted in major environmental crises. One example was when the ship “John R” stranded and broke into two pieces. Most of the oil carried by the ship was removed before the ship broke, thus inflicting only minor environmental damages. Another example was the near-incident of a 100 000 tonnes Russian oil tanker with engine failure that drifted towards the coast.
There have also been ship accidents leading to oil spills and pollution. One example is the “MS Server” grounding north-west of Bergen in January 2007. The bad weather the following days made the recovery of heavy fuel oil difficult. Around 400 tonnes of heavy fuel oil was released into the environment.

The most recent event took place in July 2009 when the Panama registered vessel “Full City” grounded south of Langesund in Southern Norway. The ship suffered severe damage to her hull, and bunker oil escaped to the sea and polluted the shorelines. Some of the affected areas were special protected areas and bird sanctuaries.

Norwegian authorities have for a long time been concerned about the risk of oil spills. One of the key measures adopted to reduce the risk of ship accidents and oil spills is to impose sailing routes for heavy ship traffic (over 5.000 GT) with high environmental risk potential farther away from the rest of the Norwegian.

Along the northern Norwegian coastline between Vardo and Røst such a more remote sailing route has been imposed. According to the Norwegian Coastal Administration (NCA), the experiences so far have been positive.

The Norwegian authorities wished to examine the consequences of prolonging the imposed route along the entire coast. The idea is that any emergencies or possible oil spills will then occur farther away from the coast, giving the authorities more time to react and enable emergency towing or oil spill response that may significantly reduce the overall environmental impact. Another effect is that possible oil spills from ship accidents will then to a greater extent evaporate before reaching the coast.

An expert group consisting of relevant stakeholders got the task to propose a route alongside the remaining coast. Their proposal will be referred to as the “proposed route” while today’s traffic pattern will be referred to as “today’s route”.

The Norwegian Coastal Administration (NCA) commissioned The Institute of Transport Economics (TØI) to conduct a risk assessment of the proposed route which led to two reports this article is based upon (Akhtar J. and V. Jean-Hansen V., 2009a).

The overall objective of this study is to produce quantitative risk analyses in order to estimate the effects of the proposed measure. The main research objective is thus to compare the accident risks and accident consequences with and without the proposed route implemented.
Risk calculations will be conducted both by use of current ship traffic volumes and by use of traffic forecasts for the year 2025.

2. Method

The approach adopted is to compare the proposed route with today’s route given traffic data of 2008 and 2025 by simulating the number of accidents in each route. The traffic information of 2008 was gathered from the universal Automatic Identification System’s (AIS) data and today’s route was made using the AIS charts which revealed the traffic patterns. An accident and risk simulation programme called MARCS, developed by DNV Technica, was used for this purpose.

We specified the following six relevant vessel types according to their potential for bunker and cargo spills:

- Chemical tanker
- Gas tanker
- Oil tanker
- Cargo ships larger than 5000 GT
- Other ships larger than 5000 GT
- All other vessels

By first defining the routes of today’s traffic we could define specific crossing sections on the routes. Data from the AIS database were utilized in order to collect information on all vessels on defined routes and the crossing sections in the year 2008. For our analyses we required AIS data of traffic volumes sorted by weight (GT and DWT), tanker type, IMO number and speed. These data were then used as input into the accident simulation program.

Today’s route have no separation schemes. However the oil transport flow in the Norwegian Sea goes from north to south and from the offshore oil platforms to the plants onshore. Hence two lanes were defined on top of each other for the tankers, one fully loaded (north-south) and one mainly empty going back towards north. The rest of the traffic was defined as constantly half loaded.

In the case of drifting or other emergency situations, collisions can occur with these structures leading to oil spill, loss of ship etc. Therefore, information about the installations at sea, vessels have to navigate past, like platforms, wind parks etc. were also inserted into the MARCS program.
When simulating the different probabilities for the proposed route, all vessel types except for “all other vessels” were moved to the proposed lanes. These vessels are not considered to have high potential environmental risk potential.

2.1 AIS data
The universal Automatic Identification System – AIS – is a ship-borne transponder that broadcasts information about the ship, the voyage, and several safety-related issues. AIS information is transmitted between vessels, from vessels to shore, or vice versa. In simple terms, AIS is a technology to make ships “visible” to each other. (Mou J.M. et.al. 2010) The coverage of the system is similar to other VHF applications, i.e. it depends on the range to the horizon from the antenna. A total of 37 stations form the AIS network in Norway. Typically the range is 45 NM from the coast. Research is currently in progress in order to increase this range by using a space-based AIS receiver which will have a range of up to 1000 NM (Eriksen et al., 2006).

2.2 The simulation
The Marine Accident Risk Calculation System (MARCS) was developed by Det Norske Veritas (DNV). The marine traffic image data used by MARCS is a representation of the actual flows of traffic within the calculation area. MARCS uses various models to calculate the probabilities and frequencies of collision and oil spills, these models are: the Collision Model, the Powered Grounding and Powered Collision Models, the Drift Grounding and Drift Collision Models, Repair Recovery Model, Recovery of Control by Emergency Tow, Recovery of Control by Anchoring, the Structural Failure Model and the Fire and Explosion Model.

The data needed by MARCS model to run a simulation are classified into four main types:
- Shipping lane data describes the movements of various marine traffic types within the study area;
- Environment data describes the conditions within the calculation area, including the location of geographical features (land, offshore structures etc.) and meteorological data (visibility, wind rose, currents and sea state);
- Internal operational data describes operational procedures and equipment installed onboard ship – such data can affect both accident frequency and accident consequence factors;
- External operational data describes factors external to the ship that can affect ship safety, such as VTMS (Vessel Traffic Management Systems), TSS (Traffic Separation Schemes), and the location and performance of emergency tugs – such data can affect both accident frequency and accident consequence factors.
In our research the shipping lane data was collected using the AIS database. The environment data and external data were provided by the NCA while the internal operational data and external operational data were provided by the DNV.

DNV obtain these data by using either worldwide data or frequency factors obtained from fault tree analysis or location specific survey data. These data are treated like a trade secret which includes:

1. The probability of a collision given an encounter;
2. The probability of a powered grounding given a ship’s course is close to the shoreline;
3. The frequency (per hour at risk) of fires or explosions.

Two simulations were run, one for today’s route, and one for the proposed route.

### 3. The routes

The AIS data was used to plot all vessels of 100 metres length or more for the period 2008-2009. 100 metres was chosen after checking the AIS files and confirming that all tankers were at least 100 metres.

The plots in figure yield the typically used routes for large vessels today. A single sailing is marked with a yellow line in today’s routes. Areas with high density are marked red. On the basis of these traffic data, today’s routes could be determined. Caution was taken to define a new lane at every shift of traffic density or angle. At every corner of these lines, a new lane was defined and the AIS data extension was made. In the plots, today’s routes are indicated by the parallel lines running along the coast. The crossing lanes are indicated as shaded grey areas. The proposed route which had identified by an expert group consisting of relevant stakeholders was assumed to not influence the crossing traffic mostly being from the platforms to the mainland. The crossing lanes were thus held equal for both simulations.

<table>
<thead>
<tr>
<th></th>
<th>Today’s route</th>
<th>Proposed route</th>
</tr>
</thead>
</table>


<table>
<thead>
<tr>
<th>Mid</th>
<th>West</th>
<th>South</th>
<th>Figure 1 Today’s route and the proposed route</th>
</tr>
</thead>
</table>

**Transport sustainability and environment**

64
4. Traffic forecast 2025

The ships sailing in the Norwegian fairways will be of many different nationalities, but the traffic is dominated by vessels linked to the offshore activities in Norwegian and Russian areas. Along the fairway outside the Norwegian coastline, ships to and from the Norwegian ports dominate the traffic.

Passenger traffic is limited and predominantly consist of cruise ships (> 5000 GT) coming from Europe and America visiting Norwegian fjords, North Cape and the coast around Svalbard. Oil and gas fields are currently being developed in the northern area of the study area, and further developments both on Russian and Norwegian fields are likely in the period 2008-2025. Most of the oil and gas produced in these areas has to be exported by ship from the fields to the markets in Europe and North America. Thus the numbers of oil and gas tankers are expected to increase.

In addition small tankers carrying fuel to the fishing fleet, oil and gas for residential heating etc. and other types of cargo ships will be travelling along the coast. The last group of ships (“all other vessels”) is dominated by fishing vessels going to the fishing fields and mostly to Norwegian fishing ports for further export by ship, road or air. Russian trawlers also deliver fish to Norwegian ports.

Our prognosis was developed by splitting Norway geographically into six parts due to differentiated traffic developments. By assessing the traffic in these six areas and by analyzing the oil and gas shipments plan from the NCA and from internationally published reports a forecast was developed. The forecast was divided into two, one for tankers and one for non tankers larger than 5000 GT.

The forecast given in figure 2 is based on expected export figures from the Shtokman field given by Frantzen B. and Bambulyak A. (2008), as well as on forecasts of the transport of oil and gas from the Norwegian oil and gas fields given by Hovi and Madslien (2008). According to these prognoses, tanker traffic is expected to increase from 2014 onward, notably gas tanker traffic. This is closely related to the planned development of the gas production from the Shtokman field.
In addition to the tanker transit traffic given in figure 2 there is also tanker traffic to and from the oil and gas fields in the North and Norwegian Sea and product tankers trafficking Norwegian refinery ports. Also an increase in the product tanker traffic going to and from Norwegian ports is expected. Traffic to and from the oil and gas fields in the Norwegian and North Sea is, however, expected to remain stable. Figure 3 presents the forecast of the total tanker traffic between the entire study areas distributed by traffic types.
For crossing lanes we have assumed a larger increase in the number of sailings with LNGs compared to oil tankers. The reason is that we expect an increase in household gas consumption compared to oil consumption, because gas is more environmentally friendly and clean. In addition, gas is easily available and cheap for the populations of the North Sea countries. Table 1 below gives the traffic volume data for today’s route for 2008 and 2025 distributed by type of ship. The numbers are aggregated per ship type by MARCS. These traffic figures are used as the basis for the risk calculations presented later in the article.

Table 1 The total traffic volume (in 1000 nautical miles) in 2008 and 2025 distributed by type of ship on today’s routes.

<table>
<thead>
<tr>
<th></th>
<th>Today’s routes 2008</th>
<th>Today’s routes 2025</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical Tanker</td>
<td>958</td>
<td>1420</td>
</tr>
<tr>
<td>Gas Tanker</td>
<td>327</td>
<td>1450</td>
</tr>
<tr>
<td>Oil Tanker</td>
<td>609</td>
<td>903</td>
</tr>
<tr>
<td>Cargo ships &gt;5000 GT</td>
<td>1250</td>
<td>1500</td>
</tr>
<tr>
<td>Other ships &gt;5000 GT</td>
<td>464</td>
<td>554</td>
</tr>
<tr>
<td>All other vessels</td>
<td>4430</td>
<td>4020</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>8040</strong></td>
<td><strong>9850</strong></td>
</tr>
</tbody>
</table>

We see from figure 3 that implementing the proposed route will increase the traffic volume of all ship types, but only marginally for “all other vessels”. The forecast for 2025 reveals a substantial increase in the traffic volume of tankers, notably gas tankers. Traffic volumes of “all other vessels” are expected to decrease in 2025, due to a decreasing number of fishing vessels. The ship traffic volumes are visualized in table 4. Table 2 shows the keys for reading these plots. Red colour indicates dense traffic with over 10 movements per day, while green, blue and grey colours indicate low levels of traffic. Yellow colour indicates traffic of 1-5 movements per day while orange indicates 5-10 movements per day. Vessels smaller than 5000 GT have been excluded from the diagrams, however they are included in the simulations.

Table 2 Key to Ship Traffic Plots

<table>
<thead>
<tr>
<th>Colour</th>
<th>Traffic frequency (ship movements per day within each location)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.05 to 0.1</td>
</tr>
<tr>
<td></td>
<td>0.1 to 0.5</td>
</tr>
<tr>
<td></td>
<td>0.5 to 1</td>
</tr>
<tr>
<td></td>
<td>1 to 5</td>
</tr>
<tr>
<td></td>
<td>5 to 10</td>
</tr>
<tr>
<td></td>
<td>&gt; 10</td>
</tr>
<tr>
<td>Year</td>
<td>Traffic plots for all traffic &gt;5000 GT</td>
</tr>
<tr>
<td>------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Today’s routes</td>
</tr>
<tr>
<td>08</td>
<td><img src="image" alt="Traffic plots for 2008" /></td>
</tr>
<tr>
<td>25</td>
<td><img src="image" alt="Traffic plots for 2025" /></td>
</tr>
</tbody>
</table>

Figure 3 Traffic plots for all vessels above 5000 GT in 2008 and 2025 on today’s route and on the proposed route. Today’s traffic (2008) is given in the top left diagram; estimated traffic on the proposed route in 2008 is given in top right diagram. Traffic forecasts for 2025 on today’s route are given in the bottom left diagram and on the proposed route in the bottom right diagram.
5. Simulation results

The accident and oil spill calculations were confined to the release of the following materials due to accidental events:
- Crude oil and refined products carried as cargo by tankers.
- Bunker fuel oil carried by all ships.

Tables below present the expected accident frequency per year given traffic volumes and ship type distributions of 2008 and 2025 for the routes travelled today and for the proposed new route.

Table 4 Expected accident frequencies per year

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Chemical Tanker</th>
<th>Gas Tanker</th>
<th>Oil Tanker</th>
<th>Cargo &gt; 5000 GT</th>
<th>Other vessels &gt;5000 GT</th>
<th>Rest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Today’s routes 2008</td>
<td>8,26</td>
<td>1,01</td>
<td>0,42</td>
<td>0,51</td>
<td>1,28</td>
<td>0,49</td>
<td>4,55</td>
</tr>
<tr>
<td>Proposed route 2008</td>
<td>7,77</td>
<td>0,70</td>
<td>0,31</td>
<td>0,48</td>
<td>1,25</td>
<td>0,46</td>
<td>4,58</td>
</tr>
<tr>
<td>Today’s routes 2025</td>
<td>10,30</td>
<td>1,51</td>
<td>1,70</td>
<td>0,77</td>
<td>1,55</td>
<td>0,60</td>
<td>4,20</td>
</tr>
<tr>
<td>Proposed route 2025</td>
<td>9,11</td>
<td>0,95</td>
<td>0,97</td>
<td>0,87</td>
<td>1,57</td>
<td>0,56</td>
<td>4,19</td>
</tr>
</tbody>
</table>

Table 5 Expected spilling accident frequencies per year

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Chemical Tanker</th>
<th>Gas Tanker</th>
<th>Oil Tanker</th>
<th>Cargo &gt; 5000 GT</th>
<th>Other vessels &gt;5000 GT</th>
<th>Rest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Today’s routes 2008</td>
<td>1,42</td>
<td>0,43</td>
<td>0,15</td>
<td>0,22</td>
<td>0,13</td>
<td>0,05</td>
<td>0,46</td>
</tr>
<tr>
<td>Proposed route 2008</td>
<td>1,23</td>
<td>0,30</td>
<td>0,11</td>
<td>0,20</td>
<td>0,12</td>
<td>0,05</td>
<td>0,46</td>
</tr>
<tr>
<td>Today’s routes 2025</td>
<td>2,13</td>
<td>0,64</td>
<td>0,53</td>
<td>0,33</td>
<td>0,16</td>
<td>0,06</td>
<td>0,42</td>
</tr>
<tr>
<td>Proposed route 2025</td>
<td>1,70</td>
<td>0,42</td>
<td>0,31</td>
<td>0,34</td>
<td>0,16</td>
<td>0,06</td>
<td>0,42</td>
</tr>
</tbody>
</table>

Table 6 Expected oil spills in tonnes per year

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Chemical Tanker</th>
<th>Gas Tanker</th>
<th>Oil Tanker</th>
<th>Cargo &gt; 5000 GT</th>
<th>Other vessels &gt;5000 GT</th>
<th>Rest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Today’s routes 2008</td>
<td>5130</td>
<td>1080</td>
<td>430</td>
<td>3530</td>
<td>58</td>
<td>6</td>
<td>27</td>
</tr>
<tr>
<td>Proposed route 2008</td>
<td>4540</td>
<td>875</td>
<td>314</td>
<td>3240</td>
<td>75</td>
<td>6</td>
<td>27</td>
</tr>
<tr>
<td>Today’s routes 2025</td>
<td>8710</td>
<td>1610</td>
<td>1670</td>
<td>5330</td>
<td>70</td>
<td>7</td>
<td>25</td>
</tr>
<tr>
<td>Proposed route 2025</td>
<td>5040</td>
<td>729</td>
<td>509</td>
<td>3740</td>
<td>52</td>
<td>4</td>
<td>13</td>
</tr>
</tbody>
</table>
We see from table 4 and 5 that the total number of accident is decreased with the proposed route. However oil tanker and huge cargo ship accident are somewhat increased for the year 2025. We see the same pattern in table 6 which show the expected oil spilling accidents. The expected number of spill accidents is decreased with the proposed route. This is true both for the total number of accidents and for accidents involving chemical tankers and gas tankers. For oil tankers there is a small increase in the expected number of oil spill accidents when the proposed route is adopted, for the year 2025. The main reason for this effect is probably that lanes in the proposed route narrows the traffic and thus increases the probability of a collision.

5.1 Expected accident frequencies and oil spill volumes for tankers
Oil tankers are of particular interest in the present study because they contribute to the vast majority of oil spills. For oil tankers the proposed route has particularly beneficial effects. One important reason for this is that grounding accidents are reduced (Akhtar, J., V. and Jean-Hansen, 2009b). Figure 5 presents the simulated oil spill volumes from oil tankers and the rest of the fleet on today’s route and on the proposed new route in 2008 and 2025.

![Figure 5](image)

Figure 5 Expected oil spill volumes from oil tankers and the rest of the fleet in tonnes, including bunkers oil on today’s route and the proposed route with traffic data for 2008 and 2025.

Figure 6 has been constructed by calculating the effects of the proposed route using table 5-7. We see that transferring ship traffic to the proposed new route gives a six per cent reduction in all accidents and a 13 % reduction in oil spill accidents and 12 % in oil spill including bunkers oil.
spill in 2008. For 2025 the numbers are 12, 20 and 42%, respectively. However, only the oil spill reductions are calculated to have a statistically significant change for 2008 and 2025.

![Figure 6 Expected effects of proposed routeing. Per cent change in all accidents, oil spill accidents and the volume of oil spills with traffic data for 2008 and 2025.](image)

### 6. Discussion and Conclusions

It is clear that the oil spill probability is reduced for all vessel types for traffic scenarios, 2008 and 2025, by adopting the proposed route. In total, the proposed routing is expected to reduce oil spills by 590 tonnes per year in 2008 and by 3670 tonnes in 2025. The main reason for this substantial reduction is that the number of groundings is reduced since the distance from the shore is increased. The reduction is particularly strong for tankers.

MARCS simulations were restricted to accidents affecting the marine environment, while accidents in port approach and port areas are not. The simulation does not take into account the possible higher alertness of the crew and easier navigation which follows by the separation of the traffic. Neither does the simulation adjust for the positive experiences derived from traffic separations schemes elsewhere in the world (IMO 2009). Accordingly it is possible that our simulation overestimates the accident frequency of oil tankers in 2025.

Another effect of the transfer of ship traffic to the proposed route is that possible oil spills will occur farther away from the coast, giving the authorities more time to react and enable
emergency towing or oil spill response that may significantly reduce the overall environmental impact.

However, when oil spills occur farther away from the coast a higher number of sites may be exposed to the spill. Experience shows nevertheless that the number of impact sites is not proportionate with the total impact area and may also be counterbalanced by the lower concentration of oil (severity of impact) on each site.

Summing up the results, it seems clear that implementing the proposed route for tankers and vessels above 5000 GT will lower the probabilities for accidents and oil spills and also reduces the volumes of potential oil spills reaching the coasts. Thus, the proposed new route will significantly reduce the environmental impacts of the shipping traffic along the Norwegian coast.

**References**


BRINGING TRANSPARENCY TO ENVIRONMENTAL IMPACTS OF TRANSPORT - LIPASTO TRANSPORT EMISSION DATABASE

Heidi Auvinen, VTT Technical Research Centre of Finland
heidi.auvinen@vtt.fi

ABSTRACT

Environmental data, such as emission factors and energy consumption figures, are needed in order to assess the environmental burdens caused by transport. Firstly, the data enables quantification of the environmental impacts, useful for example when analysing the role of transports in the life cycle of a consumer product. Secondly, environmental information is required for making comparisons, for example when managing complex logistics chains and choosing the means and mode of transport. The demand and use of environmental data concerning transport is growing, and the user group spreads all the way from individual citizens and logistics service providers to top-level decision-makers. A number of databases, calculation methods and tools have been set up, but in absence of a harmonised and standardised approach, results from different sources may be incomparable and lead to misguided conclusions.

The LIPASTO transport emission database by VTT Technical Research Centre of Finland is one of the most comprehensive sources to offer unbiased, free of charge emission factors and energy consumption figures for all four modes: road, rail, waterborne and air transport. While the main goal and focus is to cover the Finnish transport system, including international passenger and goods transport, a great part of the data applies to all Europe. The updates and further developments of the database will be done in accordance with on-going standardisation work concerning energy consumption and greenhouse gas emissions in relation to transport services (CEN/TC 320 WG10). Another initiative in the field to involve the LIPASTO database is an up-coming European Commission funded research project COFRET (Carbon footprint of freight transport). The COFRET project aims at establishing a concrete methodology to determine the carbon footprint along complex logistics chains. Furthermore, the LIPASTO emission database figures are currently being converted into a format that enables them to be incorporated into the life cycle assessment (LCA) software KCL-ECO.

This paper presents the LIPASTO unit emission database and describes how it tackles some of the problematic issues in the field. These include how to take into account the empty trips in a logistics chain, how to allocate the resulting environmental impacts and how comparisons should be made between means or modes of transport. In addition, the next steps in the development work, especially linked to on-going standardisation and research projects, will be sketched out.
1. INTRODUCTION

Environmental impacts of transport cover all changes to the environment caused by the transport system. These include both the positive and the negative aspects throughout the life cycles of the actual transports and related infrastructures, energy systems and vehicles. Emissions to air, soil and water, noise and natural resource depletion are examples of the negative side, in other words, environmental burdens. In order to manage the impacts, extensive knowledge and understanding of natural sciences and engineering, combined to information and data production, is required.

This paper discusses the operational use-phase energy consumption figures and emission factors that represent one dimension of environmental data used to assess the environmental burdens caused by transport. In other words, the discussion applies to real-world energy consumption and emission accounting and reporting regarding both passenger and freight transport. For a study choosing the focus in a different manner, for example, concentrating on vehicle design, the criteria for evaluation would be different. This paper addresses energy and emission related transport efficiency and comparisons in terms of overall performance, not the only the theoretical, technical context.

The demand for afore-described transport energy consumption and emission data is constantly growing, and the user group spreads from individual citizens, companies and logistics service providers to regional, national and international decision makers. The needs can be divided into two. Firstly, data on transport energy consumption and emissions is needed in order to measure and quantify the environmental impacts, in this case most importantly relating to climate change, air quality and fossil fuel depletion. One application is the quantification of the carbon footprint of a consumer product, a significant part of which may result from transport operations. Secondly, information on energy consumption and emissions is required for making comparisons, for example when choosing the means and mode of transport. Analogous decisions are made on a daily basis by individuals planning their daily travel routines, operators planning complex logistics chains and authorities responsible for public transport planning or infrastructure investments.

A great number of databases, calculation methods and tools have been set up to fulfil the need for environmental information on transport energy consumption and emissions. However, in absence of a harmonised and standardised approach, results from different sources are often incomparable and may thus lead to misguided conclusions. Whereas one source provides emission factors for a bus with maximum number of passengers aboard, another source takes the actual number of passengers and the empty return trips into account. Therefore completely different results can be derived even for identical cases. While both results are correct, they do, however, serve very different uses. If this goes unnoticed, misinterpretation of the results may lead to most unfavourable decisions. Currently the lack of standardised practices combined to poor understanding of the parameters behind energy consumption figures and emission factors are being addressed by on-going standardisation work and growingly transparent databases. Technical Committee CEN/TC 320 is preparing a standard that will establish a common methodology for the calculation, declaration and reporting on energy use and greenhouse gas emissions related to a transport service of goods, passengers or both. The
standard will guide on how to treat parameters such as capacity use and return trips and how to allocate the burdens. State-of-the-art databases, calculation methods and tools are following the standardisation work closely and as a first step towards comparability, some of them are now reporting and explaining their respective choices on afore-mentioned parameters in a transparent way.

This paper contributes to the discussion on transparency and comparability of energy consumption and emission factor databases by observing the field, reviewing the state-of-the-art initiatives and by presenting the LIPASTO unit emission database by VTT Technical Research Centre of Finland. The LIPASTO database is one of the most comprehensive sources to offer unbiased, free of charge energy consumption figures and emission factors for all four transport modes, and the database is maintained and improvements are made in accordance with the progress in the CEN standardisation activities. In addition, the LIPASTO system aims at full transparency and therefore all parameters behind the figures are available and explained. The purpose of this paper is to justify the need for harmonisation in terms of standardisation activities and co-operation between all stakeholders, including the owners of current databases. In other words, this review aims to contribute to and promote the ultimate long-term goal for transparent, free of charge transport energy consumption and emission data accessible to anyone, enabling both quantification of the environmental burdens and comparisons within and between different means and mode of transport.

The structure of the paper is divided into two sections, followed by concluding discussion in chapter 4. Chapter 2 explores the state-of-the-art, reviewing the on-going standardisation and harmonisation activities. Further, state-of-the-art methods and tools, related literary work and problematic areas within the topic are discussed. Chapter 3 presents a more thorough introduction to the two components of the LIPASTO transport emission database by VTT Technical Research Centre of Finland and development work around it.

2. STATE-OF-THE-ART REVIEW

2.1. Setting

Energy consumption and emission figures are needed by a wide-spread group of actors. These include individual citizens, companies, logistics service providers and decision makers at different levels. The two-fold need to first quantify the environmental impacts and then to compare the impacts between alternatives clearly calls for a common, standardised approach. Inventories on energy consumption and emissions are no longer made only at the aggregate level concerning an entire nation or a production plant. Similar calculations and reporting are being applied to individual passengers, products and transport services. Emission labels are tagged on, for example, passenger cars, public transport journeys, newspapers and food products.

Currently there is neither binding guidance nor standardised practices to define how energy consumption figures and emission factors should be constructed for passenger and freight
transport services. However, the on-going preparatory standardisation work is about to change this, and the progress made so far has already started to create a more harmonised setting. The European Committee for Standardization (CEN) has started the standardisation work in order to establish a common methodology for the calculation, declaration and reporting on energy use and greenhouse gas emissions related to a transport service of goods, passengers or both. The Working Group 10 in Technical Committee 320 (CEN/TC 320 WG10) started to the preparatory work in 2009 and the CEN Formal Vote concerning the draft of the standard should take place in 2012. The draft is expected to be available for National Standards Bodies around the end of 2012.

Current drafts of the CEN standard under preparation emphasize the inclusion of not only the operational use-phase vehicle processes but also associated energy processes (CEN, 2010). These cover exploitation, production and distribution of transport fuels. The draft also discusses the level of precision, and for example for quantification of vehicle energy consumption four levels are identified but also ranked according to preference. The main outcome with the standard when applied will be better consistency and transparency in transport energy consumption and emission reporting in a manner where energy consumption and emissions are fully allocated to the transport service payload.

2.2. Existing methods and tools

The existing methods, tools and databases for transport energy consumption and emissions consist of free-access and commercial solutions offered by companies, transport service providers, logistics operators, environmental organisations, research institutes and administrative authorities. The level of coverage and precision varies from poor to extremely detailed, but comparisons between sources can be difficult. Despite the great number of efforts and fragmentation in terms of quality, the number of established, extensive emission factor databases to cover passenger and freight transport in all modes remains moderate. The databases offering information on transport energy consumption and emissions have not been widely discussed in the academic and scientific literature. However, some examples of analyses and comparisons have been made (Mäkelä & Auvinen, 2009; Whyte, 2010; te Loo, 2009). Also, research has been done on assessing the role of use-phase energy consumption and emissions considering the entire life cycle of a transport service (Chester, 2008; Chester & Hovarth, 2009). Besides broadening the scope to include for example transport infrastructures and fuel production in addition to use-phase impacts, another approach is to broaden the time frame. Such studies explore the long-lived effects and impacts of transport emissions and the resulting climate responses (Berntsen & Fuglesvedt, 2008).

In the following, two free-access and one partly restricted state-of-the-art emission factor databases on the internet are briefly introduced. Also, one of the most prominent efforts to realise a global transport emission calculator and choices for emission factors are discussed.

VTT Technical Research Centre of Finland. The LIPASTO emission factor database by VTT is an extensive collection of energy consumption and emission figures for road, rail, waterborne and air transport, along with working machines and off-road vehicles. Both
passenger and freight transport are covered and the database is freely accessible in Finnish and in English. The factors primarily address the Finnish transport system, but most of them are applicable in the European or even the global context. The LIPASTO approach emphasises the use of actual (realised) emission figures instead of theoretical or design values. Therefore actual capacity use and empty trips are taken into consideration and embedded in the emission factors, although also the use of own (for example company specific) capacity use and empty trip data is possible and encouraged. Included are energy consumption, carbon dioxide equivalent (CO$_2$e) and emissions of carbon monoxide (CO), hydrocarbons (HC), nitrogen oxides (NOx), particle matter (PM), methane (CH$_4$), nitrous oxide (N$_2$O), sulphur dioxide (SO$_2$) and carbon dioxide (CO$_2$). (Mäkelä & Auvinen, 2009, 2010)

**Defra, Department for Environment, Food and Rural Affairs (UK).** Defra provides an extensive database of greenhouse gas emission factors in English directed to company reporting. Detailed guidance is accompanied with emission factors for passenger and freight transport and production of transport fuels, taking into account average capacity use. Although prepared for national purposes, most emission factors are applicable in the European or the global context. However, included are only greenhouse gas emissions of methane (CH$_4$), nitrous oxide (N$_2$O) and carbon dioxide (CO$_2$) along with the carbon dioxide equivalent (CO$_2$e). (AEA, 2010)

**NTM, The Network for Transport and Environment (Sweden).** The Swedish NTM has compiled a comprehensive transport emission database in Swedish and English to cover the national (passenger transport) but also the international level (freight transport). Some features, such as basic calculation tools are freely available, but the actual database can only be accessed through membership subscription. Included are energy consumption and emissions of carbon monoxide (CO), hydrocarbons (HC), nitrogen oxides (NOx), particle matter (PM), methane (CH$_4$), nitrous oxide (N$_2$O), sulphur dioxide (SO$_2$) and carbon dioxide (CO$_2$). Figures for energy consumption and carbon dioxide can be traced back to the primary source being fossil or renewable, or in case of electricity, nuclear energy. (Swahn, 2009)

**EcoTransIT World.** The Ecological Transport Information Tool (EcoTransIT), as developed by the Institute for Energy and Environmental Research (Ifeu) and the Rail Management Consultants GmbH (RMCon), is one of the most developed freight transport emission calculators with a global approach, in English. Also, a complementary calculator for passenger transport in Europe, EcoPassenger, has been developed. The effort was initiated by a group of European railway companies in the early 2000s, but the spectrum of partners has since then grown to cover all transport modes. The two calculation tools have a broad selection of vehicles from all modes, and the user is able to control and adjust variable factors for capacity use and empty return trips. Examples of the special features in the tool are the option to include energy consumption and emissions resulting from freight handling and the option to include radiative forcing index (RFI) when assessing the climatic impacts of air transport. Environmental parameters included are primary energy consumption and emissions of hydrocarbons (HC), nitrogen oxides (NOx), particle matter (PM), sulphur dioxide (SO$_2$) and carbon dioxide (CO$_2$). (IFEU Heidelberg et al., 2010)
2.3. Problem areas

In the absence of a common agreement on how to define and report energy consumption and emissions a number of problems arise. Differing practices compromise comparability at all levels and lead to misguided decisions. Typically energy consumption and emissions are underestimated and major gaps are left in the process of acknowledging the real resource use and environmental impacts. For example, if empty return journeys are ignored, a significant part of the total energy consumption and emissions could go unnoticed. Results derived taking such a shortcut suggest superior performance compared to another actor, the one that does take return trips into account and therefore allocates all resulting emissions to payload.

Even if all relevant energy use and emissions released are acknowledged, a crucial object of dispute is how to allocate them and how to define the payload, most typically passengers, freight or both. The general principle that is presumably also adopted in the CEN standard suggests that allocation should be done based on actual, useful payload. In other words, emissions should be assigned to the real physical amount of freight or number of passengers transported, not to the theoretical maximum capacity. It also means that emissions should be allocated to net freight, excluding the weight, area and volume of the packaging, container, etc. More controversial allocation matters include discussion on how to split emissions between freight and passengers when both are transported and how to split emissions between commodities transported in the same vehicle, on the same route but from and to different addresses.

The level of input data is one more factor influencing the outcome of energy consumption and energy accounting. Resource use and emissions, as well as payload in terms of passengers or freight, can be measured for a specific transport service or as an average over a given time period. Third option is to use established default values as reference. The first alternative, making use of real-time monitoring, would in many applications be the ideal solution. However, this option is not yet available in most cases. As a guideline, the most detailed data available should be used in a systematic way, and the use of data and assumptions made should be reported to ensure transparency.

3. LIPASTO TRANSPORT EMISSION DATABASE

3.1. Emission inventory database

The LIPASTO transport emission system consists of two databases. The LIPASTO emission inventory provides the annual total transport energy consumption and emission figures by mode in Finland. This calculation system consists of four mode-specific sub models, each of which is updated with new activity data on an annual basis. In addition, a calculation model for working machines and off-road vehicles has been constructed to complement the system. Sub models for road, rail and waterborne transport, as well as the model for working machines and off-road vehicles, are developed and maintained by VTT, whereas Finavia
Corporation answers for the model for air transport. The work for developing the calculation system and the four sub models started as early as 1988, and the principle source for financing has been the Finnish Ministry of Transport and Communications. The annual calculation routines and updates are currently financed by Statistics Finland, responsible also for compiling and reporting the national transport emissions as required by the European Union and the United Nations Framework Convention on Climate Change (UNFCC). In addition, a number of organisations, mainly governmental, participate in different ways to help improve the LIPASTO system and provide it with the best, accurate data input.

The main result of the annual emission inventory calculation round is the total national mode-specific energy consumption and emissions for the following compounds: carbon monoxide (CO), hydrocarbons (HC), nitrogen oxides (NOx), particle matter (PM), methane (CH₄), nitrous oxide (N₂O), sulphur dioxide (SO₂) and carbon dioxide (CO₂). Results can also be segregated to regional level, useful format for municipalities and other actors compiling local inventories. Besides in assessing the climate implications, another noteworthy use for the LIPASTO emission calculation results is air quality studies.

### 3.2. Unit emission database

The LIPASTO unit emission database was first set up as an extension to the LIPASTO emission inventory in the late 90s to answer the growing needs for vehicle-specific emission factors. Since then the LIPASTO unit emission database has established its status as an equal with the inventory part, in terms of demand. And in the international context the interest for the unit emission database has actually outgrown that of the emission inventory. The update routine for the unit emission database is currently not as regular as for the annually revised emission inventory, but updates and improvements are intended to be carried out once a year or biennially.

In the LIPASTO calculation system, the term ‘unit emission’ refers to the use-phase energy consumption figures and emission factors of a vehicle over one kilometre. All four transport modes are covered, both regarding passenger and freight transport, with the inclusion of working machines and off-road vehicles. Typically, energy consumption and emission figures are expressed as grams per transport of a given transport unit over the distance of one kilometre, that is as grams per passenger kilometre (g/pkm) or grams per tonne kilometre (g/tkm). For better transparency and in order to encouraging the use of company-specific load factors, also values per kilometre are available for most modes and vehicles. Compounds included are: carbon monoxide (CO), hydrocarbons (HC), nitrogen oxides (NOx), particle matter (PM), methane (CH₄), nitrous oxide (N₂O), sulphur dioxide (SO₂) and carbon dioxide (CO₂). Furthermore, carbon dioxide equivalent (CO₂e) figures are calculated to address to the growing concern over global warming.

The goal with the LIPASTO unit emission database is to report energy consumption and emission factors adjusted to reflect the current transport efficiency in Finland. For passenger transport this means that energy consumption and emissions are divided by the actual number of passengers, as recorded in Finland. For example, the factors for air travel take into account
the actual cabin factor in the reference year 2008. Following the same approach, also the
choices of vehicles presented in the database aim to best represent the fleet used in Finland.
For road transport, emission factors are given for each euro class but also the annual average
car or vehicle, derived from the Finnish vehicle fleet and mileage share driven. Freight
transport energy consumption and emissions are allocated to actual amount of goods
transported. For waterborne transport, remarkably detailed information on transports has been
received from ports and authorities, and all burdens are allocated to net freight (excluding the
weight of the container or the vehicle carrying the freight). Furthermore, all return trips, even
if empty, are taken into account and thus all energy consumption and emissions resulting from
transport services are allocated either to actual passengers or net freight carried.

3.3. Implications and future work

The LIPASTO unit emission database is developed and updated according to the continuous
improvement approach. This includes involvement in the European standardisation work and
co-operation with different interest groups such as transport companies, industrial sector,
authorities and developers of similar databases. One example of the co-operative efforts is the
up-coming European Commission funded research project COFRET (Carbon footprint of
freight transport). The aim with the COFRET project is to establish the framework and
methodology for determining the carbon footprint along complex logistics chains and for
implementing the carbon footprint calculations into current supply chain management
systems. The LIPASTO database will serve both as input material and point of reference or
comparison in the project.

In addition, the LIPASTO emission database is widely employed in life cycle assessment
(LCA) calculations. Therefore the energy consumption and emission figures are currently
being converted into the format to support the LCA software KCL-ECO. The commercially
available software and the extensive database for research purposes will then be even better
prepared to quantify the impacts of transport in the life cycle of, for example, a consumer
product.

4. Discussion

Environmental data on transport energy consumption and emissions is needed in order to
quantify and compare resulting environmental impacts and burdens. Carbon footprints are
calculated and reported, for example, for consumer products, freight, passengers, vehicles and
transport services. Demand for such environmental information is growing and more detailed,
transparent approaches are required.

This paper discussed the use-phase transport energy consumption and emissions factor
databases with special focus on transparency and comparability. Currently there is no single
definition or method as how transport energy consumption and emissions should be calculated
and reported. Different databases and calculation tools lead to completely different results
depending on their choice of approach. Moreover, incomplete reporting of the background
information on underlying assumptions and choices compromises the transparency and comparability, the two most fundamental qualities in assessing environmental impacts.

A number of initiatives have been launched to address the growing need for environmental information on transport and the problems related to the diverse practices in producing this information. Most importantly, the European Committee for Standardization (CEN) is preparing a standard that will establish a common methodology for the calculation, declaration and reporting on energy use and greenhouse gas emissions related to a transport service of goods, passengers or both. On the other hand, also the major actors in providing methods, tools and databases to enable transport energy consumption and emission reporting are contributing in the standardisation work and opening up the details behind the existing applications. This co-operative approach to support harmonisation should not be seen as a disadvantage in economic terms or in terms of competitiveness.

This paper gave an overview of the CEN standardisation work, three state-of-the-art transport emission databases (by VTT, Defra and NTM) and one transport emission calculator (EcoTransIT World). Further, a more detailed introduction was given to the LIPASTO transport emission database developed and maintained by VTT Technical Centre of Finland. The LIPASTO database provides energy consumption figures and emission factors for all four transport modes and working machines, and the database is being developed in accordance with the ongoing CEN standardisation. The database and all background information to ensure complete transparency are accessible on the internet, free of charge.

ACKNOWLEDGEMENTS

This paper is written based on the research work funded most importantly by the Finnish Ministry of Transport and Communications, Statistics Finland and VTT Technical Research Centre of Finland. The author wishes to extend special thanks to her colleague and the founder of the LIPASTO system Kari Mäkelä. Also, support and valuable comments from Anu Tuominen (VTT Technical Research Centre of Finland) and Rocío Cascajo (Transport Research Centre, Madrid University of Technology) while writing this paper are greatly appreciated.

REFERENCES


PHOTOCATALYTIC ROADS: FROM LAB TESTS TO REAL SCALE APPLICATIONS
ABSTRACT

Photocatalytic concrete exhibits air purifying properties. Under the action of (sun) light, a catalyst (TiO$_2$) present at the surface of the material is activated, enabling degradation of pollutants from the surroundings and transformation to less harmful products. It is a promising technique to reduce a number of air contaminants such as NO$_x$ and VOC’s, especially at sites with a high level of pollution: highly trafficked canyon streets, road tunnels, urban environment, etc. In addition, the combination with cement-based products offers some synergistic advantages, as the reaction products can be adsorbed at the surface and subsequently washed away by the rain.

However, the great potential of this emerging technology is hampered by the lack of uniform testing methods and procedures at European level (CEN) to assess, in an objective manner, the photocatalytic activity of these new building materials. For this purpose, laboratory research is undertaken at the BRRC to compare existing test methods, identify important influential parameters and draw up recommendations for future standards. Furthermore, the translation of laboratory testing towards results in situ remains critical to demonstrate the effectiveness on large scale applications. In this perspective, several trial applications have recently been initiated in Belgium to assess the “real life” behavior.

This paper gives a short overview of the photocatalytic principle and the application in concrete, as well as some main results of the laboratory research recognizing the important parameters that come into play. In addition, the implementation and preparation efforts of some recent realizations in Belgium will be presented.
1. INTRODUCTION

Emission from the transport sector has a particular impact on the overall air quality because of its rapid rate of growth: goods transport by road in Europe (EU-27) has increased by 46% (period 1995-2008), while passenger transport by road in the EU-27 has gone up by 20% and passenger transport in air by 62% in the same period (European Commission, 2010). The main emissions caused by motor traffic are nitrogen oxides (NO\textsubscript{x}), hydrocarbons (HC) and carbon monoxide (CO), accounting for respectively 58%, 50% and 75% of all such emissions (Beeldens, 2008).

These pollutants have an increasing impact on the urban air quality. In addition, photochemical reactions resulting from the action of sunlight on NO\textsubscript{2} and VOC’s (volatile organic compounds) lead to the formation of ‘photochemical smog’ and ozone, a secondary long-range pollutant, which impacts in rural areas often far from the original emission site. Acid rain is another long-range pollutant influenced by vehicle NO\textsubscript{x} emissions and resulting from the transport of NO\textsubscript{x}, oxidation in the air into NO\textsubscript{3} and finally, precipitation of nitric acid with harmful consequences for building materials (corrosion of the surface) and vegetation.

The European Directives (Council of the European Union, 1999) impose a limit to the NO\textsubscript{2} concentration in ambient air of max. 40 µg/m\textsuperscript{3} NO\textsubscript{2} (33 ppbV) averaged over 1 year and 200 µg/m\textsuperscript{3} (163 ppbV) averaged over 1 hour. These limit values gradually decreased from 50 and 250 in 2005 to the final limit in 2010.

Heterogeneous photocatalysis is a promising method for NO\textsubscript{x} abatement. In a first part of this paper, the principle of photocatalytic materials will be elaborated, followed by a description of the past laboratory research indicating important influencing factors for the purifying process. Next, an overview of the results regarding the first pilot project in Antwerp (Beeldens, 2008) is given, and finally, different applications in Belgium that have recently been started, will be discussed.

2. HETEROGENEOUS PHOTOCATALYSIS, A PROCESS FOR AIR PURIFICATION

A solution for the air pollution by traffic can be found in the treatment of the pollutants as close to the source as possible. Therefore, photocatalytically active materials can be added to the surface of pavement and building materials (Chen and Poon, 2009). In combination with (sun) light, the pollutants are oxidized due to the presence of the photocatalyst and precipitated on the surface of the material. Afterwards, they can be removed from the surface by the rain or cleaning/washing with water (see Figure 1).
Heterogeneous photocatalysis with titanium dioxide (TiO₂) as catalyst is a rapidly developing field in environmental engineering, as it has a great potential to cope with the increasing pollution. The impulse for the use of TiO₂ as photocatalyst was given by Fujishima and Honda in 1972 (Fujishima et al., 1999). They discovered the hydrolysis of water in oxygen and hydrogen in the presence of light, by means of a TiO₂-anode in a photochemical cell. In the eighties, organic pollution in water was also decomposed by adding TiO₂ and under influence of UV-light (with wave lengths lower than 387 nm). The application of TiO₂, in the photoactive crystal form anatase, as air purifying material originated in Japan in 1996 (e.g. Sopyan et al., 1996). Since then, a broad spectrum of products appeared on the market for indoor use as well as for outdoor applications. Regarding traffic emissions, it is important that the exhaust gases stay in contact with the active surface during a certain period. The geometrical situation, the speed of the traffic, the speed and direction of the wind, the temperature, all influence the final reduction rate of pollutants in situ.

In the case of concrete pavement blocks (Cassar and Pepe, 1997; Murata et al., 1996), the anatase is added to the wearing layer of the pavers which is approximately 8 mm thick. The fact that the TiO₂ is present over the whole thickness of this layer means that even if some abrasion takes place by the traffic, new TiO₂ will be present at the surface to maintain the photocatalytic activity. Another, similar application consists of using a double layered concrete with addition of TiO₂ (in the mass and/or as dispersion on the surface) to the top layer, which will be discussed later on. The use of TiO₂ in combination with cement leads to a transformation of the NOₓ into NO₃⁻, which is adsorbed at the surface due to the alkalinity of the concrete (Cassar et al., 2007). Thus, a synergetic effect is created in the presence of the cement matrix, which helps to effectively trap the reactant gases (NO and NO₂) together with the nitrate salt formed. Subsequently, the deposited nitrate will be washed away by rain.

Up till now, UV-light (in the UV-A spectrum) was necessary to activate the photocatalyst. However, recent research indicates a shift towards the visible light (Fujishima and Zhang, 2006). This means that applications in tunnels and indoor environments become more
realistic. Especially the application in tunnels is worth looking at due to the high concentration of air pollutants at these sites. One of the projects in Belgium is focusing on this subject (PhotoPAQ, 2010).

3. LABORATORY RESULTS: PARAMETER EVALUATION

Different test methods have been (and still are being) developed to determine the efficiency of photocatalytic materials towards air purification. An overview is given by Cassar et al. (2007). A distinction can be made by the type of air flow; in the flow-through method according to ISO 22197-1 (2007), the air, with a concentration of 1 ppmV of NO, passes over the sample which is illuminated by a UV-lamp with light intensity equal to 10 W/m² in the range between 300 and 400 nm, as illustrated in Figure 2. Afterwards, the NOx concentration is measured at the outlet. It is also worth to note here that within Europe actions are underway to harmonize and develop new standards for photocatalysis (CEN, 2007). Specifically concerning NOx abatement, investigations are currently being made, also at the BRRC, into a new type of mixed reactor system which could offer some advantages in the future. In any case, the test procedure used for the current results is following the existing ISO-standard.

![Figure 2: Schematic and photo of measurement set-up according to ISO 22197-1 (2007) at the BRRC.](image)

The preparation of the samples is of great importance. Due to the photocatalytic activity NO₃⁻ is deposited on the surface and covers to a certain extent the TiO₂ from the light and pollutants. Consequently, the efficiency is lowered over time, but by rinsing the surface, the initial efficiency can be restored again as also demonstrated later on. The pre-treatment of the samples in the laboratory can be important to obtain reproducible results and mainly depends on the type of base material. In the case of concrete, the release of NO and NO₂ prior to the photocatalytic reaction is limited; in the case of paints however, this can be more important. A typical test scheme according to the ISO standard is represented in Figure 3, where the following steps are applied to the sample: 0.5 hour at 1 ppmV NO-concentration, no light – 5 h exposure to an air flow of 3 l/min with 1 ppmV NO and UV-illumination – 0.5 hour with
UV-illumination and no exposure. A small increase in time of the NO\textsubscript{x} concentration is visible due to the deposit of the NO\textsubscript{3}⁻ at the surface.

![Graph showing concentration over time](image)

Figure 3: Typical result obtained in laboratory according to the standard ISO test procedure.

In the laboratory, the influence of different important test parameters affecting the photocatalytic reaction has been investigated (Beeldens, 2008) such as temperature, light intensity, relative humidity, contact time (controlled by surface area, flow velocity, height of air flow over the sample…). For instance, the effect of relative humidity of the ingoing air is illustrated in Figure 4. Clearly, the reduction of the NO\textsubscript{x} concentration in the outlet air decreases with increasing relative humidity (RH %). This has to do with the fact that the water in the atmosphere plays a role in the adhesion of the pollutants at the surface and the competition effect that can arise between water molecules and NO\textsubscript{x} in the ambient air with increasing RH. Hence, relative humidity is an important limiting factor for photocatalytic applications in humid areas like Belgium.
In general, it can be stated that the efficiency towards the reduction of NO\textsubscript{x} increases with a longer contact time (larger surface area, lower air velocity, smaller height of air flow, higher turbulence at the surface), a lower relative humidity and a higher intensity of incident light. These are the conditions at which the risk of ozone formation in summer is the largest: high temperatures, no wind and no rain. At these days, the photocatalytic reaction will be more pronounced.

4. PILOT PROJECT IN ANTWERP

An important issue is the conversion of the results obtained in the laboratory to real applications. In order to see the influence of photocatalytic pavements in “real life”, a first pilot section of 10.000m\textsuperscript{2} of photocatalytic pavement blocks was constructed in 2004-2005 on the parking lanes of a main road axe in Antwerp. Figure 5 gives a view of the parking lane, where the photocatalytic concrete pavement blocks have been applied. Only the wearing layer of the blocks contains TiO\textsubscript{2}. In spite of the fact that the surface applied on the Leien of Antwerp is quite important, one has to notice the relative small width of the photocatalytic parking lane in comparison with the total street: 2*4.5 m on a total width of 60 m.

Figure 4: Effect of relative humidity on photocatalytic efficiency.
Two different types of tests were carried out. First of all, pavement blocks were taken from the Leien after different periods of exposure. These blocks were measured in the laboratory with and without washing of the surface. The results are presented in Figure 6 and demonstrate a good durability of the efficiency towards NOx abatement. The deposition of pollutants on the surface leads to a decrease in efficiency which can be regained after washing. To check the longevity of the photocatalytic action, measurements were recently (in 2010) repeated on in situ removed paving stones, as shown in Figure 7. The results indicate that even after more than 6 years of service life, the durability of the photocatalytic pavers still persists.

![Figure 5: Separate parking lanes at the Leien of Antwerp with photocatalytic pavement blocks.](image)

![Figure 6: NOx concentration at the outflow, measured on 2 pavement blocks, before (z.w.) and after washing the surface (purple).](image)
Besides the tests in the lab, on site measurements were also carried out. Since no reference measurements without photocatalytic material (prior to the application) exist, the interpretation of these results is rather difficult. Especially the influence of traffic, wind speed, light intensity and relative humidity are playing an important role. Detailed results can be found in Beeldens (2008). In brief, the field measurements suggested a decrease in NO\textsubscript{x} concentration at the sites with photocatalytic materials, where a levelling out of the peaks is visible. In any case, precaution has to be taken with the interpretation of data since these results are momentary and limited over time. But, at least, they give an indication of the efficiency of the photocatalytic pavement materials in situ, and a basis to work on for future applications.

5. **RECENT PHOTOCATALYTIC APPLICATIONS IN BELGIUM**

Since the first application in Antwerp (2004-2005), much progress has been made within the photocatalytic research area. Newer, better and more efficient materials are constantly being developed, and action is more and more broadened also to visible light responsive materials. An example of such a material is given in Figure 8. Hence, the need still exists to develop more *in situ* applications in which the relation between the efficiency in laboratory and on site is established (*e.g.* Guerrini and Peccati, 2007; Maggos *et al.*, 2008; Gignoux *et al.*, 2010). An overview of two such recent projects in Belgium is given in this section.
Increasing and decreasing light intensity

Figure 8: NO-reduction under visible light as a function of different intensities (BRRC, 2009).

5.1 Life+-Project PhotoPAQ

The European Life+-project PhotoPAQ, *Demonstration of Photocatalytic remediation Processes on Air Quality* (PhotoPAQ, 2010), is aimed at demonstrating the usefulness of photocatalytic (road) construction materials for air purification purposes in an urban environment. Within this consortium, consisting of 8 partners coming from 5 different countries, two extensive field campaigns will be organized within Europe, of which one in Belgium. For the latter, photocatalytic cementitious materials will be applied on the side walls and roof of the Leopold II tunnel in Brussels (see Figure 9).

Figure 9: Inside view of test site within Leopold II tunnel in Brussels for PhotoPAQ project.

A test section of about 200 m in length is foreseen, which will be renovated in summer 2011. Afterwards, an intensive measurement campaign will take place in September 2011 to
rigorously assess the effect on the air pollution inside the tunnel. To this end, a dedicated UV-lighting system will also be installed inside the tunnel which can be modulated (on/off) to directly see the action of the photocatalytic walls. Concurrently, simulations of the tunnel air flow will be performed in order to model the abatement of pollutants and the effect of different influencing parameters (traffic flow, concentration profiles, ventilation…). This modeling, when validated with measurements, could provide a valuable tool for extrapolation of the findings to other sites. First results for this project are expected in beginning of 2012.

5.2 INTERREG Project ECO2PROFIT

The broad environmental sustainability project ECO2PROFIT deals with reduction of the emission of greenhouse gasses and sustainable production of energy on industrial estates in the frontier area between Flanders and Holland. To reach these goals, several tangible demonstration projects have been planned on industrial sites in Belgium and the Netherlands.

One particular project is situated on the industrial zone “Den Hoek 3” in Wijnegem (near Antwerp). Here, the regional development agency POM Antwerp is aiming to use a double layered concrete for the road construction, with recycled concrete aggregates in the bottom layer and photocatalytic materials (TiO₂) in the top layer. That way, air purifying and CO₂ reducing concrete roads can be built which are both innovating and energy efficient. For this project the BRRC was asked to set-up an elaborate testing program in the lab to help optimize the air purifying performance of the top layer, without interfering with other properties of the concrete (workability, mechanical, durability…). Furthermore, a trial section was constructed, as shown in Figure 10, to get familiar with the technique of constructing two-layered concrete.

![Figure 10: Trial section of double layered concrete with TiO₂ in the top layer on industrial zone “Den Hoek 3” in Wijnegem.](image)

For the application of photocatalytic materials in a concrete road (and in general for any other type of application) a fundamental choice can be made between: mixing in the mass (TiO₂ in cement) and/or spraying on the surface (dispersion of TiO₂). The former has the advantage of a more durable action since the TiO₂ will continuously be present, even after wearing of the
top layer. On the other hand, the initial cost will be higher (higher TiO$_2$ content, necessity for double layered concrete) and only the TiO$_2$ at the surface will be active. In contrast, dispersing at the surface of a TiO$_2$ solution will provide a more direct action, and a lower initial cost (e.g. “ordinary” cement). In this case however, the longevity of the photocatalytic action could be questioned because of loss of adhesion to the surface in time. This fundamental choice was also investigated within the research programme.

For reasons of noise reduction and comfort of the road user, it was decided to use an exposed aggregates surface finish (grain size 0/6,3) for the top layer (see right side of Figure 10). In a first phase, an optimization of the concrete composition for both layers had to be performed in terms of workability (“ideal” grading curve), mechanical properties (compression strength), and durability (resistance against frost/thaw cycles), as illustrated in Figure 11.

Subsequently, trial sections of 30 m long and 3 m wide (left side of Figure 10) were constructed to assess the feasibility of using the double layered technique in practice. In this stage, a “normal” cement was used (no mixing of TiO$_2$ in the mass) while a photocatalytic dispersion was sprayed over the exposed aggregates surface, after application of a curing compound. In addition, some first tests towards the air purifying performance of the concrete surface were made. To this end, test plates were also made on site using the concrete of the top layer and with application of the photocatalytic dispersion. After construction, these plates were kept under atmospheric conditions for 20 days, and subsequently cut to size and stored at RH = (60±2)% and T = (20±2)°C for 15 days. Finally, the samples were tested for their
photocatalytic efficiency with the set-up of Figure 2 and according to the ISO-norm. The results are depicted in Figure 12 and clearly show substantial and repeatable photocatalytic efficiency for these first test materials.

**Figure 12**: Photocatalytic efficiency obtained for test plates of trial section in Wijnegem (photocatalytic dispersion on the surface).

In order to improve the air purifying performance, further testing was undertaken in the laboratory in which the effect of different, important factors was studied:

- Effect of different materials in the mass
- Effect of different dispersions on the surface
- Influence of curing compound
- Influence of curing and/or storing conditions
- Effect of surface finishing
- Simulation of durability

First of all, it appeared that different photocatalytic materials available on the market (for mixing in the mass as well as applying on the surface) can give drastically different results regarding their air purifying performance, as shown in Figure 13.
A second important influencing factor is the curing compound, which is normally applied after exposing the aggregates at the surface, to protect the concrete against desiccation. Time of applying is approximately 24 hours after putting the concrete in place. Its effect is illustrated in Figure 14.

Apparently, the curing compound will initially inhibit the photocatalytic reaction, probably because it is shielding off the active components from the pollutants in the air. Consequently, the curing must disappear from the surface (normally after 1-2 months open to traffic) before the TiO$_2$ will reach its optimal air purifying performance. In case of a TiO$_2$ spray, this also means that it is best to apply the photocatalytic dispersion some time after the curing compound to have the best effect.
Besides the curing compound, also the storage and curing conditions of the concrete play a role, although to a lesser extent compared to the former. The effect is most pronounced in the case of absence of a curing compound, where it can be seen that more humid conditions have an adverse effect on the photocatalytic efficiency. This is related to the relative humidity conditions at the surface of the concrete and the competition effect between water and pollutants as described above. Moreover, the hardening process of the concrete will slightly differ depending on the curing conditions which could in turn affect the porosity of the surface and hence, also the photocatalytic action. This could be important in practice, because it is obviously hard to control these hardening conditions in situ.

To see the effect of surface treatment and more specifically of the exposed aggregates surface finish, a comparison among three different surfaces has been made: exposed aggregates, smooth (formwork side) and sawn surface (for one type of product). The results are depicted in Figure 15. This shows that the exposed aggregates surface performs equally well as the smooth, formwork surface, but not as good as a sawn surface. This is the result of the combined action of less cement at the surface and a higher surface porosity, two competing effects which in the end yield the final efficiency shown in Figure 15.

![Figure 15: Effect of surface treatment on photocatalytic efficiency (only one type of “less” active product in mass).](image-url)
Finally, the durability of the photocatalytic action was also tested in laboratory by simulating the possible effect of traffic and/or weathering on the surface, through brushing and washing of the samples. The influence of this action is illustrated in Figure 16, where F, D and D’ correspond to different types of photocatalytic samples (mass and/or dispersion, with or without curing).

**Figure 16: Effect of brushing and washing the samples in the lab (simulation of durability).**

The photocatalytic efficiency decreases by about 10% after the brushing/washing operation. This demonstrates once again the need to assess the durability of these photocatalytic materials *in situ* and to check to longevity of the action after several years of service life.

In conclusion, the effect of the curing compound, curing conditions and surface finish has been clearly clarified, as well as the durability of the photocatalytic action in the lab. Based on these results and the optimization of the concrete composition, a proper selection of photocatalytic materials and of application procedures could be made, for the construction of double layered, photocatalytic concrete roads on the industrial zone “Den Hoek 3” in Wijnegem. A final choice has been made to combine TiO2 in the mass of the concrete top layer with application of a curing compound on the surface, and spraying of a photocatalytic dispersion after 2-3 months. Furthermore, provisional controls of the photocatalytic efficiency in the lab and in situ, are planned to check the separate action of the two types of photoactive materials (mass and dispersion), and assess the longevity of the air purifying performance. Due to the hard winter of December 2010 in Belgium, the construction works only started in March 2011, after which the photocatalytic efficiency will be followed in time (2011-2012).
6. CONCLUSIONS AND PERSPECTIVES

The use of photocatalytic pavement materials in order to minimize the air pollution by traffic is applied more frequently on site in horizontal as well as in vertical applications, also in Belgium. Laboratory results indicate a good efficiency towards the abatement of NO$_x$ in the air by using photocatalytic materials. Also, the durability of the photocatalytic action remains intact. However, the relative humidity is an important parameter which may reduce the efficiency on site. If the RH is (too) high, the water will be adsorbed at the surface and prevent the reaction with the pollutants.

Measurements on site in the past indicated a decrease of the pollution peaks due to the presence of the photocatalyst. Repeated measurements in the laboratory on photocatalytic concrete pavement blocks confirm the efficiency over time, even after more than six years of service life. Although a reduction in efficiency is evident due to the deposition of NO$_3^-$ on the surface, the original efficiency can be regained by washing the surface.

The translation from the laboratory results to the “real” site efficiency is still a difficult factor, because of the great number of parameters involved. Hence, there is still a need for large scale projects to demonstrate the effectiveness of photocatalytic materials on site, including also other positive effects (O$_3$, VOC’s, PM…). To this purpose, two recent applications have also been started up in Belgium, which show already some promising results. Furthermore, the best results will be achieved by modeling the environment, validating the model by measurements and implementing the different parameters to assess the real life effect (Luminari, 2007; Moussiopoulos et al. 2008). One must bear in mind that photocatalytic applications are only effective in case of good contact between pollutants and the active surface. Parameters as wind, street configuration and pollution source play an important role.

REFERENCES


ACKNOWLEDGEMENTS

The author wishes to thank IWT Flanders (Institute for the Promotion of Innovation by Science and Technology in Flanders), Life+ and EFRO (European Union), and INTERREG for the (financial) support of the different projects.
Abstract

This study examines the intentions 15-year old Danes have of obtaining the driving licence and owning a car in the future, in order to identify the influencing and hindering factors contributing to their intentions. An extended model of the Theory of Planned Behaviour provides the theoretical background of the study, with an additional emphasis on environmental concern, willingness to accept limitations of the car, habit and travel socialization. Given that young Danish females have caught up with males when it comes to using the car, it is suggested that there should be no difference on the intention to obtain the driving licence. Furthermore it is believed that different factors influence the intentions of males and females even though their intentions may be expressed in the same way.

The sample consisted of 3000 Danish youngsters born in 1995 from all over Denmark. Participants filled out an internet based questionnaire after receiving information regarding the research. Data analysis using ANOVA identified the main differences between the groups, namely that there was not a significant difference between the genders or living area when it came to intention to obtain the driving licence, but there was a significant difference when it came to owning a car in the future, were males expressed stronger intention. Living area was as well significant, with those living in Copenhagen and surroundings expressing the least need to own a car. The correlation analysis of associations between the constructs showed that in many cases these varied in effect between the genders. These were the demographical variables, attitudes to the bike and current use of the bike, future expectations, values, current travel and habit. However, when it came to Perceived behavioural control (PBC), subjective norms (SN), willingness, attitudes and environmental constructs, the influence these had on intentions were similar for both genders.

The results give insight into future travel intentions of young Danish males and females; they have identified factors influencing their views and how these factors have different associations to their intentions. Therefore the study broadens the understanding of the influences behind the travel intentions of young Danes.
Keywords and topics:
Young drivers, travel behaviour intention, theory of planned behaviour, environmental concern, gender difference.
Research domain:
Transport sustainability and environment.

INTRODUCTION

The European Union member countries have set a target to reduce emissions of greenhouse gases by 20% below the 1990 level before 2020 and in addition, sustainable energy sources should be increased by 10% within the transport sector (Ministry of Climate and Energy, 2010). In Denmark the transport sector is the largest user of energy and road transport is the largest growing energy user, but overall CO₂ emissions from transport in Denmark have increased by approx. 31% between 1990 to 2007 (Danish Energy Agency, 2010). The private car fleet in Denmark has also increased by near 30% between 1994 and 2009 (Statistics Denmark, 2010), which has resulted in more journeys being made by car. According to the Danish national travel survey [Transportvaneundersøgelsen], car use as either a passenger or a driver by Danish youngsters between 16 and 19 years of age has increased alone by 7% for males and 12% for females between the years 1994-2009. Use of public transport and biking by this age group has furthermore decreased for the same period. In 2009 the genders used the car equally as much for their travel after the females caught up with the males. Females have as well increased the length travelled over the same time period. This gives indications to believe that there is a diminishing gender difference regarding car use for this age group (Sigurdardottir, 2010). However the reasons for these changes over the years have not been clearly identified and it is not known if the same factors are influencing this age group travel pattern and choices. In the RAC Foundation Report from 1995 (cited in Mackey, 1998) it is proposed that the key to understanding transport changes in society lies in greater understanding of those members of the population who are still in their formative years. Since young people are in a transitional stage between childhood and adulthood, they are in the initial stages of adopting new travel patterns and habits, and therefore their travel patterns are very much open to influence (Mackey, 1998). In the light of the CO₂ reduction goals that the Danes are obliged to fulfil, there are reasons for concern over the development that future generations will take. In order to anticipate and influence future development, it is important to understand the intentions that the youth has with regard to driving, their future expectations, attitudes and the degree of willingness to accept limitations to driving later in life. Attention has therefore increasingly been drawn to the significance of young people as the indicators of future change.
Background

In a review of Nordic research, Nordbakke and Ruud (2005) concluded that there have been too few studies’ performed on what influences young people’s travel patterns, mode choice and the tendency to obtain a driving licence. Those that have been performed were becoming outdated, with limited focus and samples. They also emphasised that habits evolve during this age and stressed the importance of further research in this area for future transport planning. Habit is an important concept in mode choice research, but Bamberg and Schmidt (2003) demonstrated that car use is often determined by habit, rather than conscious decision-making thereby supporting the view that car use is a habitual choice process that usually involves routine-shaped automatic associations between stimulus situations and habitually chosen options. Furthermore, habit is a key factor in environmentally significant behaviour (Stern, 2000). When adults manifest a strong habit for car use or car dependency, it is presumed to influence the children in the family. Baslington (2008), who described the process of travel socialization, emphasizes that the learning mechanisms associated with other aspects of social life are applicable to travel behaviour. The theoretical inference of travel socialization is that our thinking and attitudes towards transport modes are embedded in childhood. These are reinforced from multiple sources of influence and incorporated in major social institutions: the home, family, employment, access to leisure, shopping facilities and so forth (Baslington, 2008). She concludes that potential of young people to make an independent decision in the future regarding transport has already been undermined by an awareness of mainstream cultural values and priorities emphasizing cars as the main mode of transport. Research findings from UK actually indicate that the travel behaviour intentions of young people between the ages of 11 and 18 are dominated by the desire to drive and/or to own a car in the future (Line & Chatterjee, 2010; Derek Halden Consultancy (DHC), 2003; Mackey, 1998), despite the proportion of young adults (aged 17-20) with a full driving licence has decreased since the early 1990’s. This trend has though started to reverse in recent years (Department for Transport, 2010). However, little is known about what determines young people’s intentions in this context. Line & Chatterjee (2010) explored the factors that influence the travel behaviour intentions of young people ageing 11, 15 and 18-years of age. They furthermore explored this in an environmental perspective, the participant’s willingness to tackle climate change, and the degree to which this willingness influences, or has the potential to influence, their travel behaviour intentions. They concluded that although the participants know that there are other more environmentally friendly transport modes and alternative fuel vehicles available, they appear to place little value on the environment. They are unwilling to change their travel behaviour intentions in the light of climate change and they place higher value on identity, image and social recognition.

Young people are in a unique life stage, that in addition to their gender and living area, contributes to their attitudes and values they have about their travel intentions. Mackey (1998)
found that the reasons young people have for learning to drive and using a car were far more complex and deeply entrenched than suggested by the 'practical' advantages of driving. Psychological reasons appeared to play a key role, and the media, peer interactions and role models are crucial in sustaining the positive image of driving. The participants in the research clearly indicated that the car was the key to a desired way of life, that driving was inherently enjoyable, and that driving had become one of the most important indicators of adulthood.

Residential location goes hand in hand with transport access and those living in a dense area are more likely to be able to use other means of transport instead of the private car. Jensen (2006), studying young Danish urban dwellers, found that they expressed considerable scepticism towards cars in the city, motoring as a system of transport, and the associated problems. They were furthermore unimpressed by cars and could certainly do without them in the urban landscape. They were acutely aware of the problems posed by the car, yet they did not want any kind of restrictions on their own mobility, neither now nor in the future. Sandqvist (2002) also found that the attitudes of adolescents (regardless of car-ownership) indicated that they did not see having a car as an essential ingredient to living a good life and they did not ascribe status value to car-ownership. However, both these studies were of a qualitative nature with small samples from a densely populated area or city. Both authors indicate the need for a more representative research study.

Research on transport related attitudes, aid to identify which factors are influencing individuals which is an important part of transport research. In a review of transport related attitudes, Lorenc, Brunton, Oliver, Oliver, Oakley (2008) concluded that transport related attitudes (cycling and walking) differ across ages and that children, young people and parents are not homogenous groups. The differences need exploration through separate studies and differential analyses and furthermore they encourage future researchers to consistently report detailed analyses by age, sex, location and socio-economic status. Based on research with adult samples, Gardner and Abraham (2010) conclude that the relationships between environmental concerns, personal moral norms and attitudes have not been adequately researched in relation to car driving, and need further scrutiny. There are therefore grounds to believe that the same applies for young generations, given that young people’s motives have been less explored.

**Theoretical foundation and aim of study**

Many studies have employed psychological theories of attitude-behaviour relations such as the theory of planned behaviour (TPB) to predict mode choice and explore the importance of different factors for the actual behaviour. A central factor in the theory is the individual’s mediated intention to perform a given behaviour. Intentions are assumed to capture the motivational factors that influence behaviour; they are indications of how hard people are willing to try, of how much effort they plan to exert, in order to perform the behaviour. As a general
rule, the stronger the intention to engage in behaviour, the more the likelihood of it being performed increases. Importantly, a behavioural intention can only find expression in behaviour if the behaviour in question is under volitional control; that is, if the person can decide at will to perform or not perform the behaviour. Perceived behavioural control, together with behavioural intention, can be used directly to predict and explain behavioural achievement. At the most basic level of explanation, the theory postulates that behaviour is a function of salient information, or beliefs, relevant to the behaviour (Ajzen, 1991). The theory of planned behaviour implies the importance of social norms to the actual actions of an individual. Social norms are not considered to be some kind of standards that are valid for and accepted by all persons in a society, but rather as subjective norms that a person believes to exist in relation to significant others (see Figure 1).

![Theory of planned behaviour](image)

Figure 1. Theory of planned behaviour (from Ajzen, 1991).

Anable (2005) argues that since the theory has limitations and the addition of an independent measure of habit would improve the predictive capability of attitude-behaviour studies using it. TPB is used as the theoretical approach in this study, with several other independent measures in order to explore their influence on the participants’ intentions.

This study recruited 15-year old Danish youngsters with the aim to explore the main factors influencing their intentions for obtaining the driving licence and to own a car in the future. It is hypothesised that despite asking the participants about intentions of behaviours they have not yet performed self, that the participants have well adjusted intentions for the behaviour in particular, to obtain a driving licence in the future and to own a car in the future. Basing this on the concepts from TPB, young people that score high on intentions should score high on positive attitudes towards the car, be confident about their skills when it comes to driving (Perceived behavioural control) and be under the influence of society and norms (Social norms). Furthermore, it is expected that values play apart and those expressing more self-centred values are more likely to have strong intentions to drive. Current travel pattern, habit, travel socialization, demography and socio-economy are also important concepts in this analysis and are presupposed to influence intentions. Lastly, the participants are asked about willingness to
accept future limitations with regard to personal driving, and about how they picture the future with regard to travel, pollution and possible policy enforcement. It is thought that at this age the participants are not keen to imagine or preferring policies that make driving the personal car difficult and views toward the future of transport should be coherent to their actual intentions.

Gender was specially taken into account in the analysis and the study hypothesis that the there is not a significant gender difference when it comes to these intentions but that the factors influencing intentions differ between males and females. It is as well thought that the environmental factors (lifestyle, environmental concern etc.) insignificantly influence the participants’ intentions, but are more strongly expressed for females.

The results will be employed to cast light into what drives the future intentions of young people and how males and females differ when it comes to the reasons for wanting to drive and own a car, which should prove valuable to policymakers and planners. Identifying and targeting potentially modifiable psychological constructs underlying car use, is highly valuable in the development of effective and persuasive campaigns or interventions (Bartholomew, Parcel, Kok & Gottlieb, 2006).

**METHOD**

**Questionnaire**

Using the TBI model, the questionnaire included questions reflecting the following constructs: attitudes toward the car, public transport (PT), biking and walking, and were measured on a 5-point Likert scale from ‘not important’ to ‘highly important’; subjective norm (SN), which included questions referring to the perceived social pressure to pass the driving test and buy a car; and perceived behavioural control (PBC), referring to how the youngsters perceive the primary obstacles to be taking a driving test, driving in the future, resources, and difficulties with regard to future transport. Finally, two questions measured their intentions to obtain a driving license and own a car in the future. These were measured on a 5-point Likert scale from “not at all” to “highly likely”. The model was extended with several other constructs which have proved useful in explaining travel behaviour. Briefly, these were: values; including 11 items based on the Schwartz value Inventory (Schwartz, 2009) which measured four dimensions of values; openness to change (three items), conservation (three items), self-enhancement (three items) and self-transcendence (two items). Future expectations toward transport which was intended to explore the picture the youth have of the future; included 16 questions on a 5-point Likert scale ranging from ‘not at all likely’ to ‘highly likely’. Habit and travel socialization, referring to how family travel pattern influences the youngster. Three questions stood alone for habit. These were formulated in the following way; ‘I could easily manage each day without a car’, ‘my
family uses the car mostly out of habit instead of need’ and ‘I’m quite flexible when it comes to choosing transport mode for my trip’. These were measured on a 5-point Likert scale. Travel socialization was represented by parents’ use of the car to work and back, both as drivers and as passengers, measured on a 5-point Likert scale from ‘nearly never’ to ‘almost always’. Current travel was measured using two questions; how often they were transported by car to school, and to leisure time activities. These questions were measured on a 5-point Likert scale from ‘nearly never’ to ‘almost always’. There were six different environmental constructs in the survey. These were created by calculating an average score for those items that theoretically represented each construct. All questions were measured using a 5-point Likert scale. Environmental concern, environmental knowledge, perceived environmental control were all represented by two questions each, environmental moral norms was represented by one item, and environmental socialization was represented by four items. Environmental lifestyle included eight questions. Lastly, one question was included which stood alone and which measured how participants perceive cars to pollute. Willingness to accept limitations included 12 items about engaging in future restrictions to limit travel (legislation, policy, etc). Willingness to accept limitations of several kinds with regard to obtaining the driving licence and driving in the future was explored by 12 questions using a 5-point Likert scale ranging from ‘not at all acceptable’ to ‘very acceptable’. A standard set of demographic and socio-economic questions including gender, residential location, parents’ educational level, number of cars and driving licences in the home, etc. were as well included. A cognitive map of the constructs and their interrelations is shown in Figure 2.

Figure 2. The constructs in the questionnaire
Note that the environmental constructs do not only influence values but the arrow is placed there to simplify the picture. Demography and socio-economy is thought to influence all constructs and is placed on top of the other concepts to indicate this.

**Participants**

The sample consisted of 3000 individuals of both genders, born in 1995 from all over Denmark and obtained with the assistance of Statistics Denmark. This resulted in sampling 1557 males and 1468 females. The survey was completed by 892 individuals, 25.6% of males and 34.1% of females. Thirteen individuals (or parents) reported that they could not participate due to handicap or special circumstances, while seven individuals had moved and never received the letters. This gave a total response rate of approximately 30%. Participation was very similar from all over the country and ranged from 23.6% to 28.8% for males and 31% to 37.5% for females after living area.

**Data collection**

The survey was an internet based survey, using a SurveyConsole platform. The survey started on February 7\(^{th}\) 2011 and finished on February 28\(^{th}\) 2011. An introduction letter was sent out to the participants with information about the survey and its purpose, and included a link to the homepage. One reminder letter was used.

**Statistical analysis**

The variables that make up the main constructs were computed from the questions using different methods. Some stand alone as one item from one question, while others represent combined average scores of two questions or more. Lastly, exploratory factor analyses were performed for larger scales, e.g. attitudes and expectations, which subsequently were reduced to several subscales. One way-ANOVA was employed to indicate significant difference between the major constructs and gender. Pearson correlation was used to indicate the strength of association between the constructs and intentions to obtain a driving licence and own a car in the future. All statistical analyses were performed using SPSS 19. Two tailed p values were considered significant at the level of .05 or less.
RESULTS

Demography

Of the 892 that completed the survey there were 494 females (55.4%) and 398 males (44.6%). Postal numbers were divided up according to residency. Distribution of the genders by living location (divided after %) is shown in Figure 3.

![Figure 3. Gender distribution after area density](image)

Other demographical and socio-economical variables included in the study were: parents educational level, the expected educational level of the young people, bike ownership, public transport seasonal pass/ticket ownership, number of cars per home, numbers of driving licences per home, access to walking and biking paths and public transport. No gender difference was found on these variables according to a one-way ANOVA.

Questionnaire constructs and gender difference

Intentions. One-way ANOVA confirmed that there was a significant difference between the genders when it came to the intention to own a car in the future ((M=4.37, SD=0.841), F(1,889)=4.805, \( p=0.029 \)) were males expressed more stronger intention to own a car. However, there was no difference observable when it came to intention to obtain a driving licence in the future. Average scores for the genders and intentions are shown in Figure 4.
When looking into the living location of the participants, young people from Copenhagen and surroundings had not as strong intentions as their peers from other parts of the country. When comparing participants from the Copenhagen area (n=92) and East Jutland (n=94) for instance this difference is easily seen (see Figure 5).

![Figure 5. Strength of intentions after living area](image)

When comparing males intention after living location a significant difference was observed regarding intention to own a car in the future (F(8, 388)= 2,985, p=.,003), and also for females (F(8, 485)= 2,338, p=.,018).

**Attitudes.** These questions were all factor analysed together and from the analysis, six factors were employed further. These were given the names: *PT user* (seven items, $\alpha$=.781), *biker* (five items, $\alpha$=.767), *practical car user* (four items, $\alpha$=.717), *environmentalist/city* (four items, $\alpha$=.738), *would like to use PT more but needs access* (three items, $\alpha$=.665), *passionate about cars* (three items, $\alpha$=.591) and *dislikes PT* (to items, $\alpha$=.499). Despite subtle differences between
the genders on most factors, there was significant difference across four factors (see Figure 6). Males were found to score significantly higher on ‘biker’ ((M=3.57, SD=.791), F(1, 877)=5.44, p=.02) and ‘passionate about cars’ ((M=3.47, SD=.86), F(1, 871)=50.425, p=.000). Females, however, scored significantly higher on the factors ‘environmentalist/city’ ((M=3.66, SD=.66), F(1, 885)=7.599, p=.006) and ‘would like to use PT more but need access’ ((M=3.50, SD=.659), F(1, 880)=27.933, p=.02).

![Figure 6. The attitude subscales and gender difference](image)

*Figure 6. The attitude subscales and gender difference*

*PBC & SN.* Perceived behaviour control (PBC) was created by calculating an average score of seven variables that theoretically are supposed to measure this construct (α=.767). Subjective norm (SN) was created by the same method using four items (α=.662). A significant difference was observed for males on PBC ((M=3.84, SD=.572), F(1, 890)=14.671, p=.000). This can be observed in Figure 7.

![Figure 7. Gender difference after PBC and SN](image)

*Figure 7. Gender difference after PBC and SN*
Environmental constructs: Females scored significantly higher on all the environmental constructs except for environmental socialization and environmental knowledge where no difference was detected (Environmental concern (M=3.44, SD=.74), F(1, 890)=18.268, p=.000; perceived environmental control, (M=3.27, SD=.647), F(1, 890)=14.593, p=.000; environmental moral norm (M=3.77, SD=.818), F(1, 890)=26.726, p=.000; environmental lifestyle, (M=3.04, SD=.658), F(1, 890)=5.104, p=.024 and finally how cars contribute to pollution, (M=4.07, SD=.851), F(1, 890)=9.229, p=.002). This is depicted in Figure 8.

![Figure 8. Environmental constructs and gender difference](image)

Future expectations and willingness. After running an exploratory factor analysis using the Principal components method with a varimax rotation, four factors emerged. These were named as follows: Green future and better access for biking and PT (six items, α=.679), Dark future with more pollution (three items, α=.683), No cars in the city, electric cars and attitude change (three items, α=.524), Technology to the rescue and unlimited driving (two items, α=.356). Females were more concerned for the environment and scored significantly higher on Green future and better access for biking and PT (M=3.54, SD=.613), F(1, 889)=8.463, p=.004. Dark future with more pollution (M=3.58, SD=.655), F(1, 889)=7.293, p=.007. Males trusted more in technology to solve the environmental problems and expected to be able to drive without restrictions in the future (M=3.25, SD=.802), F(1, 889)=7.293, p=.029. No gender difference was observed on for the questions regarding willingness.

Values. Females scored significantly higher than males on openness to change ((M=4.2, SD=.577), F(1, 890)=25.164, p=.000.), conservation ((M=4.04, SD=.546), F(1, 890)=24.968, p=.000.), self-transcendence ((M=3.97, SD=.631), F(1, 890)=20.238, p=.000.).
Travel socialization, current travel and habit. No gender difference was observed regarding parents’ use of the car to work and back, both as drivers and as passengers. Interestingly, females reported that they were driven more often to leisure time activities than males ((M=3.17, SD=1.558), F(1, 702)=15.545, p=.000.). When it came to habit, no difference was measured but females reported to be more flexible when it came to choosing transport for their trip ((M=3.52, SD=.97), F(1, 780)=5.979, p=.015.).

Strength of association between constructs and intention

Bivariate correlation using the Pearson coefficient revealed several instances of significant difference for the constructs and subscales with regard to intentions to obtain the driving licence and to own a car in the future. Table 1 details the correlations between intentions and the constructs, divided according to (living) area of residence/locality of residence. Only constructs that yielded significance in any instance are included in the table.

Table 1. Correlations between intentions and constructs

<table>
<thead>
<tr>
<th>Demography:</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education – mother</td>
<td>-.175**</td>
<td>-.218***</td>
</tr>
<tr>
<td>Education – father</td>
<td>-.196***</td>
<td>-.198***</td>
</tr>
<tr>
<td>Education – youngsters expectations</td>
<td>.053</td>
<td>-.056</td>
</tr>
<tr>
<td>Number of cars at home</td>
<td>.172***</td>
<td>.139***</td>
</tr>
<tr>
<td>Driving licence at home</td>
<td>.188***</td>
<td>.130**</td>
</tr>
<tr>
<td>Access to PT</td>
<td>-.120**</td>
<td>-.077</td>
</tr>
<tr>
<td>Access to walking paths</td>
<td>-.096*</td>
<td>-.081</td>
</tr>
<tr>
<td>Access to biking paths</td>
<td>-.090*</td>
<td>-.064</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attitudes:</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT user</td>
<td>-.154***</td>
<td>-.148***</td>
</tr>
<tr>
<td>Biker</td>
<td>-.133**</td>
<td>-.174***</td>
</tr>
<tr>
<td>Practical car user</td>
<td>.290***</td>
<td>.372***</td>
</tr>
<tr>
<td>Environmentalist/city</td>
<td>-.156***</td>
<td>-.129**</td>
</tr>
<tr>
<td>Would like to use PT but needs access to PT</td>
<td>-.088</td>
<td>-.070</td>
</tr>
<tr>
<td>Passionate about cars</td>
<td>.233***</td>
<td>.319***</td>
</tr>
<tr>
<td>Dislikes PT</td>
<td>.091*</td>
<td>.114*</td>
</tr>
<tr>
<td>SN</td>
<td>.360***</td>
<td>.437***</td>
</tr>
<tr>
<td>PBC</td>
<td>.549***</td>
<td>.530***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Environmental structures:</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environmental lifestyle</td>
<td>-.197***</td>
<td>-.196***</td>
</tr>
<tr>
<td>Environmental concern</td>
<td>-.027</td>
<td>-.063</td>
</tr>
<tr>
<td>Perceived environmental control</td>
<td>-.088*</td>
<td>-.106*</td>
</tr>
<tr>
<td>Environmental socialization</td>
<td>-.026</td>
<td>-.034</td>
</tr>
<tr>
<td>Environmental knowledge</td>
<td>-.032</td>
<td>-.095*</td>
</tr>
<tr>
<td>Cars pollute</td>
<td>-.096*</td>
<td>-.127**</td>
</tr>
</tbody>
</table>
Parents’ education had significant negative relationship to females’ intention but not for males. Number of cars and number of driving licences in the household had positive relationship to females intention but for males the fewer the licences are in the household the stronger their intentions are. The SN and PBC scales yielded the highest correlation to intention for both genders and in the same direction. All attitude subscales, except for ‘biker’, had the same impact on the genders intentions. Females are more environmentally friendly in their thought and act but despite this, the association is the same for the genders when it comes to the environmental constructs and correlation to intentions. Willingness was negatively correlated to intention, meaning that the stronger the intentions to obtain a driving licence and own a car were the more negative the view toward limitations. The genders vary with regard to their future expectations and those females that see the future in darker light, with more pollution and traffic, feel less strongly about their intentions. Of the value subscales, self-enhancement had the strongest positive relationship to intentions but this relationship was considerably stronger for males than females. Females express less car dependency and which is in turn negatively correlated to intentions. They also report being driven more often to leisure time activities which in turn had significant relationship with intentions.

**DISCUSSION**

This study is based on data from 892 individuals which constituted a 30% share of the original sample. The low response rate can partly be explained by the fact that the survey was open only for three weeks and during this period there was a one week school winter holiday. This could have influenced participation negatively. Nevertheless, the response rate yielded enough power for analysis to be performed.
Gender was specially analysed in this analysis. As previously concluded by Lorenc et al. (2008) gender and location has a major influence on the views of children, youth and parents about different modes of transport. As hypothesized, this study showed that there are no differences in the intentions to obtain the driving licence between the genders, but males are more determined to own a car in the future. Furthermore, young people living outside the Copenhagen metropolitan area express stronger intentions to own a car in the future but no difference was detected for obtaining a driving licence. The results thereby point to the importance living area has on future travel intentions. From these results it is clear that young Danes feel very strongly about their intentions to take the driving licence test and own a car in the future. Of the whole sample, only four individuals reported no intention to take the driving licence test, while only six reported that they had no intention of owning a car. These results indicate that the youngsters see the driving licence as common thing and perhaps take it for granted. It could be a cultural phenomenon, a passage to adulthood or to a certain field of freedom which they believe is their right to explore and fulfil, even though they will not become a car owner. Thereby they do not take suggested restrictions to their ‘soon to be experienced freedom’ to lightly.

This study suggested that the reasons for the youngsters’ intentions differed according to gender. This is reflected in Table 1, which indicated that correlation between several constructs differed in strength and direction after gender. These were the demographical variables, attitudes to the bike and current use of the bike, future expectations, values, current travel and habit. These constructs need to be further scrutinized in order to explore the précis differences between the genders. However, when it came to PBC, SN, willingness, attitudes and environmental constructs, the influence these have on intentions are quite similar to both genders.

The questionnaire employed in this study was created after inspiration from mode choice research on adult samples and qualitative and quantitative studies on young people on the subject, and composed to reflect every construct in a proper way. However the wording of individual questions and the creation of subscales can always be a matter of a debate.

Furthermore it is important to evaluate if there is a more efficient way of exploring the young peoples future travel intentions, and if future studies should adopt another direction altogether. Since transport behaviour is a complex act, there are almost endless factors to take into account. However, by planning a detailed study with an effective theoretical basis, there is always some knowledge to be obtained about the current act, which could provide useful for other studies. Therefore this study has laid a foundation for further analysis regarding young peoples future travel intentions. There are more analysis to be performed on this data, firstly a causal relationship between constructs needs to be defined according to the theoretical background. This would indicate where the environmental constructs have the greatest influence and how these relate to intention. A regression analysis could help establish which constructs have a predictive value with regard to intentions, while cluster analysis could identify different
segments of youngsters according to their attitudes and other constructs. In this way it would be possible to identify different stereotypes while knowing more about how the influential factors and reasons for the youngster’s intentions vary even though they express the same intention. This study has provided insight into the future intentions of Danish youngsters and how several constructs influence these intentions. Despite being only 15 years old, they have established very strong intentions. The constructs in this study are by no means exhaustive with regards to influential factors or even barriers, but can be used to clarify which factors are more important than others. This is important for policymakers and planners of future transport.

REFERENCES


Very seriously injured: In-depth investigation of road accident characteristics
and medical consequences in Germany
1 INTRODUCTION

According to the German federal road accident statistics, over the last 15 years the decrease in the number of seriously injured road casualties has been of smaller magnitude than the decrease in fatalities (DESTATIS, 2010). From 1995 to 2009, an average annual decrease of 5.5% for fatalities compares to a decrease of 4.1% for seriously injured. In 2008, 70,466 seriously injured casualties resulted in approximately 7.8 billion Euros macroeconomic costs (fatalities: 4.6 billion Euros; Höhnscheid and Straube, 2010). Therefore, the absolute numbers as well as the financial outcome indicate a significant need for action.

Despite this, up to date relatively little is known about the medical consequences for seriously injured road casualties. This is partly due to a rather wide definition of this patient group: In the federal road accident statistics a person is considered as seriously injured if he or she is treated in a hospital directly after the accident and stays there for at least 24 hours. Casualties that die within 30 days after the accident are considered as fatalities while all injured cases that do not meet the above mentioned 24h-criterion are considered as slightly injured. Therefore, especially the group of seriously injured encompasses a wide variety of actual injury severities (Lank et al., 2009b). Next to patients who stay in the hospital to be monitored for only one night, we find cases with most severe injuries that undergo various surgeries and lengthy treatment within the intensive care unit. After leaving the hospital these patients might undergo long term rehabilitation, suffer of chronic pain or even lifelong disabilities. It is therefore a challenge for road safety research to establish concepts that allow a nationwide in-depth analysis of road accidents including seriously and very seriously injured casualties.

Up to now, in Germany a decent number of locally and temporally restricted studies have been conducted, many of them conducted or issued by the German Federal Highway Research Institute (BASt) (Auerbach, 2008; Busch, 1994; Höhnscheid, Lippard, Bartz, 2005; Lank et al., 2009a, b; Lefering, 2009a; Lefering, 2010; Malczyk et al., 2010; Otte and Jänsch, 2008; Sellei et al., 2011).

In the present paper - after shortly introducing the often applied medical scoring systems for the classification of injury severity - the results of three studies commissioned by BASt are summarized (Lefering, 2009a; Lefering, 2010; Otte and Jänsch, 2008). These studies use various data sources in order to shed a light on the total number of very seriously injured, as well as on the relative development of this number over the past years. Also the hypothesis is tested, whether a possible shift from former fatalities to survivors with very serious injuries can be in part attributed to an improvement in medical care. The studies further give insight into the injury patterns that result from severe road accidents depending on used mode of transport. It is not the objective of the present paper to present each study in large detail, but to summarize the current state of knowledge for very serious road casualties in Germany and make the according findings accessible to the international road safety community.
MEDICAL DEFINITION OF INJURY SEVERITY

To account for a wide variety of injury patterns and injury severities, it is important to record the respective information at the level of single body parts (Malczyk, 2010). The most widely used severity scoring system is the Abbreviated Injury Scale (AIS), which was first published in 1969 in order to determine the severity of single injuries based on its survivability (States, 1969). The latest update of the AIS was published by the American Association for Automotive Medicine (AAAM) in 2008 (Gennarelli and Wodzin, 2008).

In order to be able to present a number of casualties exceeding a certain integrated level of injury severity, the most commonly used measure is the Injury Severity Score (Di Bartolomeo et al., 2010). It takes into account the three most severe single AIS-scores. By squaring the respective AIS-scores before addition, a higher weight is attributed to the more severe injuries. The ISS correlates with mortality, morbidity and hospitalization time after trauma (ISS: Baker et al, 1974; Copes et al. 1988). Table 1 summarizes the derivation of ISS and Maximum AIS scores (MAIS) on the basis of the AIS scoring system.

Up to date, no commonly used definition of a very seriously injured casualty exists. Therefore in the studies presented below, different thresholds based on ISS and MAIS were applied.

Table 1: Scoring systems for the classification of injury severity.

<table>
<thead>
<tr>
<th>AIS - Abbreviated Injury Scale</th>
<th>Classification of severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis</td>
<td>For each injured body part a severity score from 1 to 6 is attributed:</td>
</tr>
<tr>
<td>The body is divided into six ISS body regions:</td>
<td>minor 1</td>
</tr>
<tr>
<td>• Head or neck</td>
<td>• moderate 2</td>
</tr>
<tr>
<td>• Face</td>
<td>• serious 3</td>
</tr>
<tr>
<td>• Chest</td>
<td>• severe 4</td>
</tr>
<tr>
<td>• Abdomen or pelvic contents</td>
<td>• critical 5</td>
</tr>
<tr>
<td>• Extremeties or pelvic girdle</td>
<td>• maximal (not treatable) 6</td>
</tr>
<tr>
<td>• External</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ISS - Injury Severity Score</th>
<th>MAIS - Maximum AIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Each AIS-score of three most severely injured body regions is <strong>squared</strong>. These three scores are subsequently <strong>added</strong>.</td>
<td>Highest single AIS-score of the injured patient.</td>
</tr>
</tbody>
</table>
2 STUDIES

ANNUAL NUMBER OF VERY SERIOUSLY INJURED ROAD CASUALTIES

The goal of this study conducted by the Institute of Research in the Operative Medicine (IFOM) was to present an estimate of the annual number of very seriously injured road casualties in Germany (Lefering, 2010). Since the federal statistics lack detailed medical information, data from the TraumaRegistry of the German Society for Trauma Surgery (TR-DGU) were used in order to estimate the figures for the entire country of Germany. The TR-DGU is a documentation of every accident victim that arrives at a participating hospital in a vital state and who is treated in the intensive care unit. The registry was established in 1993 as a quality management tool for the voluntarily participating hospitals and as a data basis for scientific analysis. Currently, including all injury causes, more than 6,000 cases are documented each year, representing more than 20 % of all very seriously injured patients in Germany (Kühne et al., 2006). For each patient a total of about 100 variables are assessed including a classification of injuries according to the AIS. Table 2 indicates the number of cases available in the TraumaRegistry for the years 1999 to 2008 depending on varying criteria of injury severity. In the present study the number of very severe road casualties was estimated using three different approximation methods. These estimates were calculated based on all casualties that were classified with an ISS of at least 16 (ISS-16) - a definition which has been repeatedly used to classify a major trauma (Copes et al. 1988). The conversion factor presented in Table 2 can be used to infer estimations based on other applicable definitions of injury severity. The definitions MAIS-3 and MAIS-4 (MAIS≥3 and MAIS≥4) have been reported by Otte, Haasper and Krettek (2006) and are as well under discussion as a criterion for very seriously injured road casualties. Further, ISS-9 (ISS≥9) was added since it includes all MAIS-3 patients but also considers patients with multiple AIS-2 injuries in various body regions. Intensive care in contrast might apply as a pragmatic definition that can be easily assessed. Finally, in this study polytrauma as the most conservative criterion was defined as at least one injury of AIS-4 combined with at least one further injury of AIS-3.

Table 2: Number of cases from the TraumaRegistry for the years 1999 to 2008 for different definitions of “very seriously injured”. The conversion factor can be used to calculate estimations based on other definitions than ISS-16.

<table>
<thead>
<tr>
<th>Definition of “Very Seriously Injured”</th>
<th>ISS-9</th>
<th>MAIS-3</th>
<th>Intensive Care</th>
<th>ISS-16</th>
<th>MAIS-4</th>
<th>Poly-trauma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of cases</td>
<td>17,425</td>
<td>16,932</td>
<td>16,560</td>
<td>13,910</td>
<td>11,189</td>
<td>9,213</td>
</tr>
<tr>
<td>amongst 30-days-survivors</td>
<td>15,219</td>
<td>14,728</td>
<td>14,373</td>
<td>11,757</td>
<td>9,114</td>
<td>7,356</td>
</tr>
<tr>
<td>Conversion Factor (ISS-16)</td>
<td>1.29</td>
<td>1.25</td>
<td>1.22</td>
<td>1.00</td>
<td>0.78</td>
<td>0.63</td>
</tr>
</tbody>
</table>
In the first estimation method, the number of very seriously injured in the TraumaRegistry was compared to the number of seriously injured in the federal accident statistics for five selected regions. Since regions were chosen where hospitals registered in the TraumaRegistry covered a maximum of the respective region, it was possible to infer the relative proportion of very seriously injured with respect to all seriously injured traffic casualties within that region. According to this analysis 8-10 % of all seriously injured road casualties can be classified as very seriously injured. Therefore, inferring from the absolute number of seriously injured in Germany, a total annual number of 7,500 to 9,500 very seriously injured casualties can be derived.

Within the second estimation method, based on the data available in the TraumaRegistry the average number of very seriously injured was calculated for three levels of hospital size (supraregional, regional, local). These approximations were extrapolated to the total number of hospitals of this particular size in Germany. It was revealed that the three predefined hospital sizes treat 30.2, 11.5 and 3.3 cases per year respectively. Considering the total of 874 hospitals of all three sizes in Germany, this resulted in an annual estimate of 6,800 to 10,400 very seriously injured casualties.

The third method was based on findings from national as well as international studies on the proportion of road fatalities that die before arriving at the hospital in relation to those that die after arriving at the hospital. Therefore, data from the TraumaRegistry indicating the proportion of patients deceasing after arriving in a hospital could be used in conjunction with the total number of road fatalities from the federal statistics, in order to estimate the total number of very seriously injured casualties. It was shown that for every patient that dies within the hospital there are 6.3 patients that survive in a very seriously injured condition. Assuming that 25-40 % of all road fatalities (Chiara et al., 2010; Liener et al., 2004; Maegel et al., 2007; Malczyk, 2010) die in a hospital, the annual number of 5,595 road fatalities (mean of 2002-2008) results in an estimation of annually 8,800 to 14,000 very seriously injured road casualties.

In conclusion, the presented estimation methods lead to the result that Germany faces approximately 10,000 very seriously injured casualties annually. If the “ISS-16 criterion” is changed to “treatment in the intensive care unit” the estimated number rises to 12,500 cases. It has to be kept in mind that there remains a large uncertainty about this number since the results of the three estimation methods cover a span from 7,000 to 14,000 casualties.
ANNUAL TREND OF VERY SERIOUSLY INJURED ROAD CASUALTIES

For the analysis of the development of the number of very seriously injured casualties, again data from the TraumaRegistry were used (Lefering, 2009a). In order to be able to access a somewhat larger amount of data, the criterion was set to “ISS-9 plus admission to the intensive care unit”. For the years 1997 to 2006 this definition was applicable to data of 11,362 road casualties from 67 participating hospitals. Due to the fact that the number of participating hospitals increased significantly over the ten year time span, an approach using relative instead of absolute figures was used.

The central research question to be answered was, whether the number of very seriously injured road casualties changed over the last ten years in a comparable manner to the figures from the official road accident statistics. Figure 1 indicates the relative deviation of the number of very seriously injured road casualties from the mean number of the entire time span under investigation. Over the ten years the numbers deviate by less than 10 % from the expected mean and no clear trend can be observed. In contrast, for the number of road fatalities as well as seriously injured an almost linear decrease can be observed. This pattern indicates that indeed there is no obvious trend for the number of very seriously injured road casualties. Therefore, the clearly positive development in the federal accident statistics for the years 1997 to 2006 is not represented here. In fact the number of very seriously injured road casualties seems to be unchanged.

Figure 1: Development of very seriously injured road casualties in comparison to the casualties of the federal accident statistics (fatalities, seriously injured). The average value of the ten year time-span is assumed as the expectancy value (0 %).
These findings raise the question whether the observed pattern can be explained by a reduction of mortality; that means by a shift of patients that formerly died who now survive the accidents with very severe injuries. This was investigated by applying established methods of mortality prediction to the patient collective described. Therefore, for every year from 1997 to 2006 the Standardized Mortality Ratio was calculated by dividing the observed mortality by the expected mortality rate that was operationalized by the Revised Injury Severity Classification (RISC; Lefering, 2009b). Values above 1 indicate more fatalities than expected, values below 1 indicate less fatalities than expected.

8,821 patients allowed for the calculation of the RISC-Score, which represented 74% of the total patient collective. Figure 2 indicates that the SMR did not significantly change between the years 1997 to 2004, while in comparison it decreased significantly for the years 2005 and 2006. This means, for those years more patients survived comparable injury severities and injury patterns than the years before. Regarding the official statistics, those surviving patients should reduce the number of fatalities and at the same time lead to an increase in the number of very seriously injured. The difference between the observed and the predicted mortality indicates that the absolute magnitude of the effect is about 5%. Therefore, it can be concluded that in 2005 and 2006, out of 100 patients that had formerly died in succession of a road accident five survived in very seriously injured state.

![Figure 2: Standardized Mortality Ratio (SMR) for years 1997-2006. Error bars indicate the 95% confidence interval.](image-url)
The third research question arising was whether there are any changes in the number of very seriously injured with respect to the mode of transport. Figure 3 indicates how the very seriously injured patients distribute across the modes of transport for the years 1997 to 2006. On average, 53.8% of those are car or truck occupants, 32.7% are two-wheelers while 13.5% are pedestrians. Within the two-wheelers, who were differentiable from 2002 on, there appear twice as many powered-two-wheelers as compared to cyclists. When closer investigating the annual trend, it can be concluded that there is a slight relative increase for vulnerable road users (from about 40 to 50%), while the relative number of car and truck occupants has consequently decreased.

![Figure 3: Relative proportion of very seriously injured road casualties depending on mode of transport.](image)

**TRANSPORT MODES AND INJURY PATTERNS**

For the investigation of injury patterns two sources of data were assessed. First, data of 7,245 patients from the TraumaRegistry for the years 2002 to 2006 were analyzed with respect to the mode of transport (Lefering, 2009a). Figure 4 indicates how often each of six body regions presented (head, thorax, abdomen, spine, upper extremities, lower extremities) was injured within the patient group. In order to exclude injuries that were of only minor severity and therefore of lower significance, only injuries that were at least classified as moderate (AIS≥2) were included. Pedestrians and cyclists show the highest proportion of head injuries (>70%) while powered two-wheelers with 45% show the lowest rate of head injuries. The thorax represents the
second most frequent injury site in total, with car-/truck occupants and powered two-wheelers showing the highest proportion for this injury site (each >60%). Pedestrians and cyclists “only” show a proportion of less than 50%. In total, injuries of the abdomen are only half as common as head- or thorax-injuries. Here, also the car and truck occupants show the highest rate with 35%, while only 16% of the cyclists suffer of abdomen injuries. Spine injuries in contrast are most prevalent in powered two-wheelers while they are least dominant in pedestrians (<16%). Overall the upper extremities are injured in 40% of the cases, with powered two-wheelers forming the most vulnerable patient group (51%). Injuries of the lower extremities at overall 58% are again very common, with pedestrians being the most vulnerable road user group (68%). Together with head and thorax injuries, injuries of the lower extremities form the typical injury pattern of a very seriously injured road casualty.

![Figure 4: Relative frequencies of injuries of certain body parts with regard to mode of transport (TraumaRegistry, 2002-2006).](image)

As a second approach, data from the German In-Depth-Accident-Study (GIDAS) were analyzed for injury patterns and mode of transport (Otte and Jänsch, 2008). In this data base about 2,000 traffic accidents with personal damage occurring in the cities of Dresden and Hanover are collected every year since 1999 including detailed technical as well as medical data. Analysis on basis of GIDAS can be considered as representative for all accidents including personal damage that are being recorded by the police within the respective cities. Representativeness for the entire country of Germany is not necessarily met and has to be investigated a posteriori (Brühning, Otte and Pastor, 2005; Pfeiffer and Schmidt, 2007).

The analysis presented here comprises data from 4,044 seriously injured road casualties. Of these, 207 were classified as very seriously injured applying the criteria of “ISS≥21 and MAIS≥3” or “admission to the intensive care unit”. In this study seven body regions were differentiated (head, neck, thorax, abdomen, upper extremities, pelvis, legs). Car and truck occupants showed a high proportion of head and thorax injuries while for truck occupants also
injuries of arms and legs were very common. For powered two-wheelers injuries of the extremities dominate the observed pattern. Further, for the very seriously injured head and thorax injuries show a high frequency. Seriously and very seriously injured cyclists who did not wear a helmet during their accident as well as pedestrians most frequently suffer from head injuries. In addition, injuries of thorax and extremities are very common for these groups.

When comparing the injury patterns of very seriously injured with those of “only” seriously injured, one of the most significant result is a higher rate of head injuries for the very seriously injured. This is of major practical relevance since head injuries carry a high risk of long term brain damage.

Despite minor variations in methodology, both the TraumaRegistry-study as well as the GIDAS-study come to very similar results concerning the injury patterns with regard to mode of transport. Car- and truck drivers consistently show a high proportion of head and thorax injuries, while in motorcyclists injuries mainly affect thorax and legs. Very seriously injured cyclists are most vulnerable for head injuries, while pedestrians next to head injuries also show a very high proportion of leg injuries.
3 CONCLUSIONS AND OUTLOOK

The reviewed studies suggest that assuming a definition of “ISS≥16” the annual number of very seriously injured road casualties in Germany can be estimated at around 10,000 cases. For the years 1999 to 2008, this corresponds to 10 to 15% of all casualties that were considered as seriously injured according to the federal statistics. Meanwhile, the trend indicating a clear reduction in the number of fatalities as well as the number of seriously injured could not be observed for the very seriously injured casualties. An analysis of predicted versus observed mortality revealed that the unchanged number of very seriously injured casualties is likely to be at least in part caused by an improvement of intensive medical care: Casualties that formerly died due to injuries acquired in the accident now survive in a very seriously injured state. Further, there seems to be a slight trend towards an increase in the proportion of very seriously injured vulnerable road users - meaning powered two-wheelers, cyclists and pedestrians - while the proportion of car and truck occupants decreased. This trend is also observable for the severely injured in the federal accident statistics (DESTATIS, 2010). Finally, the observed injury patterns reflect the demand situation for design of road safety measures for various road user groups.

One shortcoming of the first study is, that depending on the applied criterion as well as the estimation method used, the resulting estimates vary quite substantially because a large amount of uncertainties has to be taken into account. In the future, these uncertainties might be substantially reduced by the implementation of TraumaNetworks by the German Society for Trauma Surgery that will become part of an obligatory quality management of all German trauma centers. As a result, detailed information about injury patterns and severities will be available for all trauma patients in Germany. As a short term solution for the assessment of the number of very seriously injured the use of the definition “admission to the intensive care unit” is being discussed, because this could easily be assessed by police officers by contacting the hospital. The feasibility and validity of this definition is currently under investigation within a pilot project in the city of Cologne. First results are expected for 2012.

A further shortcoming of the reviewed studies is the descriptive nature of the data as well as inconsistencies with respect to the applied definitions for very severe injuries. In order to make study results more comparable, it will be helpful to apply more standardized procedures. So far the data lack the possibility to draw inferential conclusions based on detailed information on accident characteristics (i.e. causation, vehicle parameters etc.). In order to be able to attribute certain injury patterns to characteristics of the accident on a large scale basis, the linkage of accident data that is recorded by the police and usually aggregated in the federal accident statistics with medical data is necessary. This would also allow new insights into the long known problem of underreporting of road casualties (Derriks and Maks, 2007; Ward, Lyons, and Thoreau, 2006).

To evaluate the possibility of linking accident data with medical data from the emergency
services, Auerbach (2008) conducted a feasibility study for one German county in the year 2006. On the basis of 514 accidents reported by the police and 854 road casualties being treated by the emergency services it was possible to establish links for 549 patients by using the key variables date, time and location of the accident. Therefore, it was revealed that linkage of accident data with existing medical data sources is well possible without adding new information to either of the data sets. Due to the small number of cases, this study did not allow for detailed analysis of injury severity and injury patterns, but exemplary analysis revealed similar results as the studies based on data from the TraumaRegistry or GIDAS.

Unfortunately, data from the emergency services is not only less precise than the more elaborative diagnostics at a trauma center but it is also not available for the entire country of Germany in a consistent manner. This is why in a first step this approach is not suited to achieve a nation-wide database on medical consequences of road accidents. GIDAS on the other hand is capable of very thorough in-depth analysis but at the same time is restricted to two German cities. The application of the same elaborated methodology to the entire country of Germany is not economically feasible. Data from the TraumaRegistry in contrast is available nation-wide and will come close to cover all trauma patients once the above introduced TraumaNetworks are established.

No matter what definition might be chosen for the official accident statistics, it is important to coordinate these activities on an international level. This will enable to carry on international benchmarking which is established within and outside the EU in form of the CARE database (Community database on Accidents on the Roads in Europe) by the European Commission and activities of the IRTAD (International Traffic Safety Data and Analysis) Group established by the OECD Road Transport Research Program. Currently, an IRTAD-report is in preparation that will summarize the definitions applied in the participating countries and will present approaches for the linkage of hospital and police data.
4 REFERENCES


Höhnscheid, K.-J., Lippard, D. and Bartz, R. (2005). Development of the number of most severely injured victims of road traffic accidents in Germany. BASI Info Leaflets, 05/05, Bergisch Gladbach.


MEASURES FOR IMPROVING ROAD SAFETY ON RURAL ROADS IN GERMANY - (“AOSI”)
1 BACKGROUND

The road network is one of the most important prerequisites of a country's successful and efficient economy. During the last 40 years the traffic volume in the countries of the European Union increased year after year. The downside of this development is that there are still more than one million accidents causing more than 34,000 fatalities on European roads each year (European Commission, 2010). In Germany the majority of road accidents happen on roads in urban areas. But the distribution of fatalities shows that the accident severity on rural roads is twice as high as on roads in urban areas (Figure 1).

![Figure 1: Distribution of accidents and fatalities by location (own chart based on Federal Statistical Office, 2010)](image)

Against this background the need for action to reduce the number of accidents and the accident severity especially on rural roads becomes evident. Due to the fact that there are already more than 167,000 km of rural roads in Germany it is obvious that a significant improvement of traffic safety on rural roads can only be achieved by measures which are suitable for existing roads.

For this reason the Federal Highway Research Institute (BASt) established a task force called “AOSI” (abbreviation for “Außerortsstraßensicherheit”, i.e. rural road safety) in 1996 in order to improve road safety on existing rural roads in Germany by short-term up to medium-term measures. At this time around 9,000 lives were lost as a consequence of road accidents with a similar distribution to the current one as displayed in Figure 1. The “AOSI” task force was composed of experts from the Federal Ministry of Transport, Building and Urban Development (BMVBS), Federal Highway Research Institute (BASt), local road authorities of different federal states, universities and members of the German Insurance Association (GDV) to cover a broad field of competence.
2 APPROACH

In the early stage of the project the main task was to analyse the location of accidents on the rural road network, the distribution of the different types of accidents, e.g. driving accident (i.e. driver loses control of the vehicle without influence of other road users), overtaking accident, and their severity.

Based on a selection criteria of road sections with an accident rate of more than two severe accidents per kilometre within 3 years (i.e. one accident above the conspicuity threshold (FGSV, 2003)) two main factors contributing to high accident severity were identified. There was one type of road where inappropriate velocities lead to accidents due to a loss of control of the vehicle (driving accident). The probability of these accidents increased with a decreasing steadiness of the alignment in combination with bad weather and/or surface conditions. Their severity was influenced by the type and distance of obstacles in the roadside. There was a second type of road, where especially unsafe overtaking manoeuvres relating to a misjudgement of sight distance, speed of oncoming vehicles or possible acceleration lead to accidents (accident in longitudinal direction).

In order to reduce the likelihood of these types of accidents the project focussed on the enforcement of speed limits by using speed cameras (short-term measure – Figure 2) and the safeguarding of overtaking manoeuvres by constructing additional passing lanes. On sections adjacent to passing lanes with only one lane per direction overtaking bans were imposed (medium-term measure – Figure 3). This layout should prevent overtaking manoeuvres using the opposing lane to increase road safety along the whole road section.

![Figure 2: Speed enforcement](image1)
![Figure 3: Passing lanes](image2)

Based on the accident distribution analysis both types of measures were expected to have an extremely positive overall impact on road safety.

After finishing the selection process, there were 10 out of more than 50 rural road sections showing a high number of severe accidents which were chosen for this trial by the “AOSI” task force.
2.1 Speed camera enforcement

On five of the chosen test sections, where inappropriate speeds were a main contributing factor for accidents, speed cameras were installed with spacing from 500 m to up to 2.500 m. Speed cameras were placed in sections where the horizontal alignment did not meet the driver's expectation to drive at a constant speed. For this reason speed cameras were installed ahead of unexpected sharp bends or along long straights to prevent breaching of the speed limit. In total, around 65 km of the rural road network were controlled by 44 speed cameras. Each test section was equipped with induction loops to record the number of vehicles distinguished according to vehicle type, location specific time- and date-stamp, and velocity of each vehicle (see Figure 4 and Figure 5). With this procedure information about the traffic flow, the traffic composition and the vehicles' velocities can be gathered irrespective of weather conditions and time of day or night. On sections controlled by speed cameras induction loops where installed at speed camera monitored cross-sections and cross-sections without any monitoring.

![Figure 4: Location of speed cameras and induction loops](image)

In addition pursuit runs with a specially equipped car were carried out along the road section to record the velocity of the car driving in front of it. In order to avoid changes in driving behaviour the measuring equipment in the recording car was invisible for the road users. As a result, velocity information of more than 50 cars per direction recorded every 2 meters enabled displaying the speeding behaviour of car drivers along a continuous road section. Measurements were taken before and after the implementation of each measure to enable a before/after comparison and to investigate their effectiveness. Pursuit runs, carried out on speed camera equipped sections, were conducted twice after the implementation with a gap of about one year in between. Special traffic signs were used to inform the driver of the speed camera surveillance on the approaching road section. Speed cameras were announced by a sign showing “speed camera control” with a supplement of the distance. This procedure should contribute to higher acceptance of the installed speed cameras.
2.2 Passing lane and overtaking regulations

On the other five roads, where unsafe overtaking manoeuvres were a main contributing factor to accident occurrence, short passing lanes with lengths from 600 m to up to 1.200 m were built to safeguard overtaking processes. These lengths were chosen in deviation from current German road design guidelines, which specify lengths from 800 m up to 2.000 m for passing lanes. This research project was to show if short passing lane sections of 600 m can contribute to higher road safety and what their influence would be on traffic flow. The passing lane lengths and their number as well as the type of transition (critical and non-critical) depended on the local circumstances at the relevant road sections.

Induction loops were installed to record traffic and velocity data. The loops were located at the beginning, in the middle and at the end of a passing lane as well as in the section where overtaking was prohibited (Figure 5).

Road users were informed by special traffic signs about the distance to the next three-lane section in sections with only one lane in each direction and about the remaining length of the passing lane in the two-lane direction of a three-lane section.

For a time period of three years before and after the implementation of the road safety improving measures, traffic accidents were monitored. The accident data were provided by the local police authorities and contained information of date and time, the number of people involved, the accident severity as well as the manner of impact, and weather conditions. With these data the effectiveness in terms of changes in the number of accidents, the circumstances and e.g. accident rates were investigated.

In addition, road users were interviewed on routes with speed camera systems along the road and on routes with passing lanes. They were asked their opinion towards the measures and their effectiveness. Moreover there were questions about how the measures influenced their own driving behaviour and about the acceptance level of the system. This survey should represent the opinion of regular road users who are directly affected by the measures.
3 RESULTS

3.1 Speed camera enforcement

Before speed cameras were installed, the relevant sections were characterised by an excessive speeding behaviour. The percentage of drivers exceeding the speed limit was associated with the level of the speed limit before the start of the project. It was found that road sections with a lower speed limit were showing higher rates of speed limit violations than sections without a speed limit (100 kph) (Figure 6).

![Figure 6: Distribution of velocities at cross sections with different speed limit (“before” condition)](image)

After the implementation of speed cameras high velocities were rapidly reduced. On all test sections the 85th percentile speed (85% of all drivers do not exceed this velocity) almost decreased to the level of the legal speed limit, above all at cross sections with speed camera enforcement (Figure 7). One year after the speed camera implementation velocities went down by as much as 20 kph (85th percentile). This reduction in speed could be observed over almost the whole monitored road section (Figure 8). Speeding at camera controlled cross sections decreased to 5 %. At cross sections without speed cameras the reduction in speed was less obvious. Past research on this topic found similar results (see Meewes, 1992 and Andersson, G. / Larsson, J., 2005). Moreover, the research revealed that the level of speed reduction also depended on the level of speed before camera implementation.

![Figure 7: Distribution of velocities at monitored cross-sections](image)
Two years after the speed camera implementation, however, the second measurement showed a development toward more homogeneous velocities over the whole investigated road section. While driving speed did not significantly change any further at monitored cross-sections, there was a tendency to minor velocities on sections in between speed cameras. The percentage of excessive speed was reduced, too. It seems that people adapted to the new situation and noticed that it is safer and more relaxing to drive at the level of permitted speed on the speed camera enforced road sections.

![Figure 8: Pursuit run result](image)

Traffic safety increased significantly after stationary speed cameras were installed. The general spacing used in between speed cameras was 2,500 m. The development towards a homogeneous driving behaviour also contributed to a reduction of overtaking accidents and environmental pollution. High differences in speed went down so that the need for overtaking decreased, too. Road users who got used to the situation appear to have adjusted their driving habits by going slower at all times which saves energy and prevents noise and thus has a positive environmental impact.

The decreasing level of speed on camera enforced road sections had a considerable influence on the number of accidents and their severity. On some road sections, a reduction in the number of accidents with serious injury of up to 51% was achieved (Figure 9). In addition the total number of accidents was decreasing, too. There was a noticeable shift from accidents with a very high severity to a low accident severity after the implementation of the measure. Especially driving accidents and overtaking accidents were reduced. This is an obvious example for the direct dependency of speed and accident severity.
Figure 9: Number of accidents on sections with speed enforcement.

Beside the reduction of the total number of accidents, especially driving accidents (type 1) and accidents of vehicles moving laterally in the same direction (often accidents while overtaking) went down (Figure 10). The magnitude of decrease also depends on the alignment. Road sections with long straight s were often burdened with overtaking accidents (type 6). Road sections with unexpected design of bends (non-relation design) were burdened with a lot of serious driving accidents. Both types of accidents were significantly reduced due to the implementation of speed camera enforcement.

Figure 10: Types of accidents

The attitude towards the acceptance of speed enforcement differed depending on the geographical region and on the publication in the local media. For road sections in a region with many hidden speed cameras outside the scope of this research the acceptance was not as good as for other test road sections where hidden speed cameras were not used as frequently. Overall the acceptance of deploying speed cameras on high risk rural roads ranged from 75 % to 95 %. On the other side, around 65 % generally supported speed enforcement as a good tool to improve road safety on hazardous roads where speeding is the major influencing factor for accidents.
One problem which still exists are ignorant motorcyclists. Because in Germany speeding drivers caught by automated speed control must be identified from the licence plate and the picture taken (which is positively impossible due to compulsory helmet wearing), there are motorcyclists who purposely ignore the legal speed limit, thus sometimes provoking accidents with high severity.

3.2 Passing lane and overtaking regulations

On all five test sections with additional passing lanes the evaluation of velocities showed different results. The prohibition of overtaking in sections in between passing lanes leads to an adjustment of velocities in dependence of the slowest vehicle class. As a result, velocities at the beginning of a passing lane are slower than in “before”-condition (one lane in each direction without overtaking restrictions) at the same location. Half way through the passing lane section velocities are slightly higher than “before” caused by overtaking manoeuvres in the first part of the passing lane. For passing lanes itself, however, very high speeds of up to 125 kph (85th percentile) were measured. Such high levels of speed are independent of the total length of the passing lanes, i.e. it did not matter whether the passing lane had a length of about 600 m or 1200 m. In total, the highest speeds were measured in passing lane sections of 750 m to 900 m, the lowest velocities in sections of 600 m to 750 m.

Another result of the trial was that passing lanes with lengths of 600 m to up to 1200 m, deployed on sections with an AADT (annual average daily traffic) of around 5,000 to up to 10,000 vehicles a day, were long enough to dissolve platoons (a queue, consisting of at least two vehicles driving one behind the other in a certain distance) formed in sections where overtaking was prohibited (Figure 11).

After the installation of passing lanes the number of detected platoons at the beginning of the passing section was much higher than before. This fact can be attributed to the overtaking restriction ahead of the passing lane section. Moreover, the average platoon length was slightly shorter than before. At the end of the passing lane section there was a significant reduction in the amount of platoons. Especially the frequency of platoons longer than three...
vehicles was reduced. The specific trend showed that the dissolving and restructuring of platoons was also supported by short passing lane sections. The installation of passing lanes had no significant influence on travel time along the investigated road section. To get an effect on travel time, longer distances have to be designed with the used road layout. It is rather a positive effect on safe driving and overtaking because road users are informed about the distance to the next safe passing opportunity where they can pass regardless of oncoming traffic. Even though, there is a problem of speeding on passing lanes, the accident situation on sections with these additional passing lanes nevertheless changed for the better. Although, the level of road safety improvement was very different between all test sections, the overall results were showing a considerable development towards less severe accidents as well as lower total accident numbers. Formerly, the majority of crashes were accidents in longitudinal direction (accidents between vehicles moving along in carriageway). These accidents accounted for serious head-on crashes with many fatalities. By creating additional passing lanes, the risk of having such an accident was significantly reduced on all test sections (Figure 12).

There were many differences between all test tracks concerning the number of driving accidents and accidents at junctions. The development depended significantly on the local situation. Originally, test sections were chosen because of their overall poor road safety. Depending on the local situation and on the accident occurrence of many head-on crashes passing lanes were added. The sections in between passing lanes mostly remained in their original condition. On some test tracks many accidents happened in these sections for example because of bad surface conditions in combination with bad alignment. This accounted for an increasing number of driving accidents on some test tracks after the implementation of measures. Another reason, especially for driving accidents, could be identified as the end of passing lanes in combination with slippery road surfaces caused by wintry conditions. There was one test track where a number of accidents occurred under
these circumstances. Therefore, special attention must be paid to these particular areas during winter maintenance.

A survey among road users verified the suitability of this kind of road layout (passing lanes combined with overtaking restrictions in 1+1 sections). The overall majority accepted and recommended the measure. On roads with a high load of heavy vehicles in combination with only one or two passing lanes in a 10 km section, however, there was a lower acceptance of short passing lanes.

4 CONCLUSION

Speeding and dangerous overtaking manoeuvres are main reasons for a high number of accidents with fatalities and seriously injured on rural roads. The deployment of lined speed camera supervision on road sections with a very high accident severity remarkably improved road safety. The number of accidents with serious injury was reduced by up to 51 %. This success was mainly based on the enforcement of the legal speed limit. Average speed reductions of up to 10 kph (85th-percentile speed) were achieved along the whole test sections. Speed camera enforcement did not only influence the amount of driving accidents. On some test sections which had been chosen for speed camera enforcement, overtaking accidents due to high differences in driving speeds were the main factor for poor road safety. As a consequence of the speed limit enforcement differences in driving speeds decreased as well as the need for risky overtaking manoeuvres. Both consequences contributed to higher road safety.

A second approach concentrated on safeguarding overtaking. By constructing passing lane sections the need for using the opposing traffic lane during overtaking manoeuvres was avoided. Hence this measure was very effective in the reduction of head-on crashes. Accidents with personal injury and serious property damage went down by up to 64 % (general trend considered). It should, however, be mentioned that there were test tracks with just a shift in accident types, i.e. from overtaking accidents to other types of accidents, without a reduction in the total number of accidents. Changes in the number and type of accidents strongly depended on the road layout along the whole stretch. In two-lane sections where a poor horizontal or vertical alignment was not improved while passing lanes were constructed, the accident situation did not change for the better. Accidents that happened in two-lane sections with poor horizontal alignment were mainly driving accidents which cannot be prevented by restrictions on overtaking.

Although, overtaking was only allowed in passing lane sections with lane lengths from 600 m to up to 1.200 m there was no negative influence on traffic flow. The driving speed in passing lanes exceeded the speed limit but without negative influence on traffic safety. However, it is highly probable that such high overtaking speeds were also driven in before-conditions (one
lane in each direction without overtaking restrictions) but with the additional risk of having to use the lane of the opposing traffic in the passing process. That is why head-on crashes were so severe during this period. Moreover, advantages in travel time could not be measured because pursuit runs had not been conducted. It can be assumed there is no real advantage in travel time because the test sections are too short for significant time savings. The restriction on overtaking is accepted if its length does not exceed 4 km. On roads with high volumes of heavy goods vehicles the length between passing lanes should be shorter for reason of acceptance. In both cases indication signs are important to inform road users about the distance to the next passing lane section to reduce overtaking pressure.

Both, short-term and medium-term measures were very effective in improving road safety on rural roads where accidents are caused by inappropriately high speed or unsafe overtaking manoeuvres.

Future research should focus on the design and furniture of junctions along roads equipped with additional passing lanes and their influence on traffic flow and road safety. Moreover, longer road stretches with the tested road layout need to be investigated to measure its influence on travel time.

REFERENCES


German Road and Transportation Research Association (FGSV) (2003). Merkblatt für die Auswertung von Straßenverkehrsunfällen, Teil 1: Führen und Auswerten von Straßenverkehrsunfällen, Köln

Older drivers’ self-regulation in traffic

INTRODUCTION

The car is the safest and most convenient means of transportation for most older adults (OECD, 2001). Although some older adults report benefits from driving cessation such as reduced expenses and decreased stress and report that they have learned to adapt to their non-driver status over time (Rudman et al., 2006) the majority of the reported consequences of driving cessation are negative the most common being loss of independence (Rudman et al., 2006; Taylor & Tripodes, 2001). In addition, Taylor and Tripodes (2001) found that driving cessation was not associated with increase in the use of other modes of transportation - indicating loss of mobility.

Independent mobility is important for older adults’ quality of life; functionality and ability to be active members of the society (Farquhar, 1995; Avlund et al., 2004; Hakamies-Blomqvist, 2003). Consequently, loss of mobility is associated with greater need for support in daily living and increased risk of institutionalisation which poses financial costs to the society. Therefore loss of independent mobility is not only a tragedy at the personal level, it has costs at the societal level as well (Hakamies-Blomqvist, 2003). Thus it is a priority to ensure older adults independent mobility for as long as possible.

Yet research has found that many older female drivers stop driving even though they are still fit to drive (Kostyniuk & Molnar, 2008; Siren et al., 2004). Hence, Siren et al. (2004) argue that these women do not stop driving for health reasons but for financial, social, and psychological reasons. These women may be suffering the negative consequences of driving cessation for no reason with both personal and societal costs. It is therefore important to explore how to avoid that particularly female drivers stop driving prematurely.

Self-regulation of driving

Donorñio et al. (2008a) define self-regulation of driving as a strategy to continue to drive safely. Thus, self-regulation may be regarded as a means to prolong the period one can drive.
On the other hand self-regulation can also be viewed as the process leading to driving cessation and thereby the beginning of the end of driving. Either way, knowledge about self-regulation may be the key to design interventions to prevent premature driving cessation. Therefore it is important to accumulate a throughout understanding of the process of self-regulation of driving and gender differences in this.

A large amount of research has already been conducted within this area. To a large extent it has been identified how older drivers self-regulate their driving. Older drivers tend to reduce their overall amount of driving (Charlton et al., 2006; Ruechel & Mann, 2005; Raitanen et al., 2003), generally reduce their speed (Charlton et al., 2006; Ruechel & Mann, 2005) and avoid driving in difficult driving situations such as driving when it is dark (D’Ambrosio et al., 2008; Molnar & Eby, 2008; Baldock et al., 2006; Charlton et al., 2006; Ruechel & Mann, 2005; Anstey & Smith, 2003; Raitanen et al., 2003; Rimmö & Hakamies-Blomqvist, 2002; Ball et al., 1998; Holland & Rabbitt, 1992), rush hour and dense traffic (Baldock et al., 2006; Raitanen et al., 2003; Rimmö & Hakamies-Blomqvist, 2002; Ball et al., 1998), avoid complex intersections (Charlton et al., 2006; Holland & Rabbitt, 1992), motorways (Ball et al., 1998), left hand turns (Baldock et al., 2006; Ball et al., 1998), avoid driving when roads are slippery (Rimmö & Hakamies-Blomqvist, 2002), long distances (D’Ambrosio et al., 2008; Anstey & Smith, 2003; Raitanen et al., 2003), and unfamiliar areas (Donorfio et al., 2008; Molnar & Eby, 2008; Anstey & Smith, 2003).

As it is commonly challenging driving situations that are avoided, self-regulation has been regarded as a compensation strategy to make up for lost functionality. Consequently research has studied the association between functional decline and self-regulation and it appears that there is indeed an association, older drivers with reduced vision, visual attention, and cognitive functioning tend to self-regulate their driving more than healthy older drivers (Ross et al., 2009; Charlton et al., 2006; Ball et al., 1998; Holland & Rabbitt, 1992).

Even though functional decline is associated with self-regulation of driving it does not seem to, exclusively, be a compensation strategy, research has found that other factors than functional decline and health affect self-regulation. The Australian survey study by Charlton et al (2006), where 656 drivers aged 55 or older completed a phone interview, found that change in employment status such as retirement, the presence of other drivers in the household, confidence in own driving, having been in an accident, gender, and age were all related to self-regulation of driving. In fact the study by Kostyniuk & Molnar (2008) found that gender had a larger effect on self-regulation than age and functional status.

**Self-regulation of driving and gender**
Several studies have found that men self-regulate less than women (D’ambrosio et al., 2008; Molnar & Eby, 2008; Charlton et al., 2006; Rimmö & Hakamies-Blomqvist, 2002; Hakamies-Blomqvist & Wahlström, 1998). Not all studies have found gender differences though. Raitanen et al. (2003) did not find any gender differences in self-regulation, they looked at frequency of driving, mileage, situations avoided, and reasons given for reduced driving. However, women were underrepresented in the study and only urban areas were included. Ross et al. (2009) found that being female was predictor of reduced driving space (distance from home) and driving frequency, but not driving avoidance (avoiding driving in difficult traffic situations).

Studies have found that drivers generally tend to self-regulate more as they age (Ross et al., 2009; Donorfio et al., 2008; Charlton et al., 2006; Rimmö & Hakamies-Blomqvist, 2002). However, Rimmö & Hakamies-Blomqvist (2002) found that 53.6% of the female drivers refrained from diving under certain driving conditions compared to 29.7% of the males in the age group 55 – 64 years. Indicating that females self-regulate their driving from a younger age than males which again may be linked to the premature driving cessation found in female drivers as they begin the fading out of their driving too early.

Not surprisingly, it does not seem to be the biological differences per se that causes the gender differences found in self-regulation. Rimmö & Hakamies-Blomqvist (2002) suggest that lack of driving experience lead to less confidence which again leads to premature driving cessation among women. Indeed a link has been found between confidence and avoidance (Ballock et al., 2006; Charlton et al., 2006) and in the study by D’Ambrosio et al. (2008) women both showed lower confidence levels and self-regulated more than the men. Marottoli & Richardson (1998) found that confidence was related to driving frequency and mileage. In their study men were more likely to drive in risky situations, however, in the situations where both men and women drove, they were equally confident. Related to age Donorfio et al. (2008) found a decline in confidence as people age which may be part of the explanation for the increase in self-regulation found with age. Thus, confidence seems to be an important factor influencing self-regulation. On these grounds it appears that female drivers are in a vicious circle where they self-regulate their driving at a younger age which leads to decrease in driving experience and routine, which again leads to decline in confidence leading to the premature driving cessation. However, what starts this process? Is it lack of confidence? Lack of routine? Something else?

It is important to explore the older drivers own explanations for regulating their driving. What motivates female and male drivers to regulate their driving and does this change with increasing age? Are interventions aimed to increase the confidence of drivers enough to prevent premature driving cessation or are other measures needed and perhaps different measures needed for different age groups of male and female drivers?
Older drivers own explanations for self-regulation of driving

The reasons older drivers give for the self-regulation of driving depend on the type of self-regulatory behaviour (Charlton et al., 2006). The most common reason given for reduced overall driving was changes in lifestyle such as retirement and moving house, less than 20% pointed towards health problems or age related changes. Avoidance of dense traffic was also not related to functional level but rather to personal preferences, many simply reported that they did not enjoy driving in heavy traffic. Avoidance of merging with other traffic was explained with personal preference and discomfort. The reason they generally drove slower was that they now put more effort into obeying the traffic rules. On the other hand when they avoided intersections they reported safety concerns and visual problems as the reason. Likewise avoidance of rain was explained with safety concerns and finally, visual problems were often the explanation for avoiding driving when it was dark (Charlton et al., 2006).

Raitanen et al. (2003) explored reasons given for reduced driving in three countries; Germany, Finland, and Italy. The most common reason selected was “being able to reach everything without a car” followed by health problems. In Finland economic reasons were also often selected. For all three countries the option “other reasons” was selected most often. In Finland they analysed these other reasons and found that the most common reason given was retirement and not having to drive to work any longer. In fact, this reason turned out to be the most common of all reasons included in the study.

A qualitative study by Siren & Kjær (2011) found that the older drivers constructed their self-regulatory behaviour as a way to control the external risk factors posed by other road users’ risky driving. The self-regulatory behaviour was presented as reflecting driving skill and experience and thus an internal asset. There appears to be gender and age differences in the balance between internal and external reasons for self-regulation. Donorfio et al. (2008a), likewise using focus group interviews, found that females saw driving “as an intrinsic skill that changed with age due to decline in ability” whereas the males saw driving as a “static skill within a changing driving environment”. They also noted that self-regulation changed with increasing age. It begins as a method to deal with external factors such as other drivers’ bad habits, gradually self-regulation strategies to compensate for internal factors such as decline in driving ability or confidence are likewise employed.

Consequently there are reasons to believe that there are gender differences in the motivations for the various types of self-regulation and that these vary between different age groups of older drivers. As this will effect which focus interventions, to prevent premature driving cessation, aught to have, more research is needed exploring differences between subgroups of older drivers in their motivations for regulating their driving.

This paper will present some of the results of a questionnaire survey which was conducted as part of a larger project on older drivers’ self-regulation of driving. The purpose of this paper is to gain a deeper understanding of older drivers’ motivations for regulating their driving and
explore differences between male and female drivers in different age groups. In particular the paper will focus on internal versus external reasons for self-regulation of driving and explore gender and age differences.

**METHODS**

**Research design**

The research design applied in this study was a survey using structured phone interviews based on a questionnaire.

**Participants**

The participants consisted of 888 older active drivers aged 75 – 95 years, mean age was 82 years. 443 (49.9%) were females and 445 (50.1 %) males. The inclusion criteria for the study were age 75 or more and driving minimum once a month.

**Materials**

For the purpose of this study a questionnaire was developed. Decisions on which items to include was based on the literature in the area as well as two focus group interviews conducted in another part of this project. Relevant parts of the questionnaire are described below.

*Background information.* Included age, gender, marital status, driving frequency, and yearly mileage. A list of 20 symptoms and illnesses was taken from Siren (2005) and used as an objective measure of health status. In addition, questions related to cognitive difficulties and vision were included in order to get an indication of the participants self-rated cognitive functioning and vision. Finally, the participants were asked to rate their overall health.

*Changes in driving exposure.* The participants were asked to indicate if they drove more, the same or less compared to 15 years ago both in regards to driving frequency and mileage.

*Avoidance of traffic situations.* Whether the following traffic situations were avoided when possible: driving on motorway, intersections without traffic lights, roundabouts, heavy traffic,
times and places with many bikes, long distances, unknown routes, unknown areas, when
dark, when slippery, driving fast, left turns, overtaking, driving when feeling tired, driving
when feeling unwell, listening to radio when driving, and finally having a conversation while
driving.

_Reasons for reduced driving._ If the participants had indicated that they had reduced their
overall driving they were presented with a list of 22 possible reasons for this and asked to
indicate how much each applied to them. (Please see appendix for list of possible reasons)

_Reasons for avoidance of traffic situations._ If participants avoided diving on motorways, time
and places with many bikes, when dark, left turns, and when feeling unwell, they were again
presented with a list of 9-13 possible reasons for this, and asked to indicate how much each
reason applied to them. (Please see appendix for list of possible reasons for avoiding the
respective traffic situations)

**Procedure**

A sample consisting of all drivers license holders aged 75 or older was extracted from the
Danish Drivers license database. This sample consisted of a total of 125 334 license holders. Out of these a random sample of 1650 license holders was withdrawn. After excluding cases
where a phone number was not obtained or incorrect, the person not able to participate in an
interview because of e.g. language barriers or illness, a total of 1332 persons were contacted.
of these 11% were not reached (after 20 attempts) and 19% refused to participate, thus
leaving a sample of 930 (70%). Of these further 42 (3%) drove too seldom to fulfil the criteria
for inclusion. In total 888 persons completed the interview.

A letter presenting the study and informing that they would be contacted by phone and asked
to participate in a phone interview had been mailed out prior to the phone interviews. The
interviews took on average 30 – 35 minutes to complete.

**RESULTS**

The 888 participants were divided into three age groups. In the age group 75 – 79 there was
291 participants, 147 (50.5%) females and 144 (49.5%) males. In the age group 80 – 84 there
was 301 participants 150 (49.8%) females and 151 (50.2 %) males, and finely in the 85+ age
group 296 participants, 146 (49.3 %) females and 150 (50.7 % males).
Background information

Marital status. More females (41.5%, 59.3%, 74.7%) than males (16.7%, 25.2%, 34.7%) in all age groups were widowed and the percentage of both females and males who were widowed were higher the older the age groups.

Objective health. The sum of the 20 symptoms and illnesses was calculated as an objective health score for each participant ranging from 0 to 20, the higher the score the poorer health. The participants were generally healthy. The females scored 1.6, 1.7, and 1.9 in the respective age groups and the males 1.7, 1.8, and 2.1.

Subjective health. The participants were asked to report their subjective overall health on a four point Likert type scale ranging from 1 = “very good”, 2 = “good”, 3 = “less good”, and 4 = “poor”. In the analysis “very good” and “good” was combined to indicate a positive rating of subjective health and “less good” and “poor” was combined to indicate a negative rating of overall subjective health. The majority of both females (96.6%, 94.0%, 92.5%) and males (93.8%, 93.4%, 95.3%) in all age groups gave a positive rating of their overall health.

Self-rated cognitive functioning. In order to get an indication of self-rated cognitive functioning the participants were presented with seven examples of cognitive difficulties for example “difficulties concentrating” or “reacting too slowly”. They were required to indicate how often they experienced these difficulties on a five point Likert type scale ranging from 1 = “very rarely”, 2 = “rarely”, 3 = “neither rarely nor often”, 4 = “often”, and finally 5 = ”very often”. In the analysis the sum of these were calculated in order to obtain a self-rated cognitive functioning score. The score ranges from 7 – 35 and again the higher score the poorer cognitive functioning. The Cronbach’s Alpha was calculated for the scale to test for internal reliability. The results showed that the scale has good internal reliability Cronbach’s Alpha = 0.78.

The participants generally had a low score indicating that they only experienced few problems with their cognitive functioning. Females 10.3, 10.6, 10.6, and males 9.3, 10.5, 11.0.

Self-rated vision. Finally, the participants were asked to indicate how often they experienced problems with their vision on the same five point Likert type scale as the one used for cognitive difficulties. The mean score was calculated for each group. The score range from 1 – 5 where a higher score indicates poorer self-rated vision. Again both females (1.3, 1.3, 1.4) and males (1.3, 1.4, 1.4) in all age groups had low scores indicating that they did not experience much problems with their vision.
Driving and changes in driving

The participants were asked to report how often they drove.

Table 1 Driving frequency of females and males in the respective age groups

<table>
<thead>
<tr>
<th>Driving frequency</th>
<th>75 – 79 years</th>
<th>80 – 84 years</th>
<th>85 +</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Females (n=147)</td>
<td>Males (n=144)</td>
<td>Females (n=150)</td>
</tr>
<tr>
<td>Everyday</td>
<td>35.4%</td>
<td>44.4%</td>
<td>27.3%</td>
</tr>
<tr>
<td>Several times a week</td>
<td>53.1%</td>
<td>47.9%</td>
<td>60.7%</td>
</tr>
<tr>
<td>Once a week or less</td>
<td>11.6%</td>
<td>7.6%</td>
<td>12.0%</td>
</tr>
</tbody>
</table>

As can be seen from table 1 the participants, particularly the males, drove relatively frequently.

In addition, they were asked to report their yearly mileage in kilometres. Some outliers were identified using a cut off Z score of 3. In the age group 75 – 79 years two outliers were identified among the females and one among the males. In the 80 – 84 years age group there were three among the females and one among the males. Finally, in the 85+ group there were three female outliers and two males. Also, some participants had not answered this questions which further explains the reduced N seen in figure 1.

The females in all age groups reported lower yearly mileage. Also, the yearly mileage for the males is lower the higher the age group whereas this trend is not evident among the females. (See figure 1)
The participants were asked to indicate whether they drove less, the same, or more, both in regards to driving frequency and amount, compared to 15 years ago.

Figure 2 Percentage of female and male participants in the respective age groups who reported driving less frequent compared to 15 years ago.

Figure 3 Percentage of the female and male participants in the respective age groups who reported driving less (mileage) compared to 15 years ago.

As is evident from figure 2 and 3 more participants seem to have reduced their yearly mileage than their frequency of driving. For the females, the percentage who report reducing both the amount and frequency of driving increases both from the youngest to the middle age group and from the middle to the oldest age group. For the males, it is only between the youngest and the middle age group there is an increase in the percentage reporting having reduced the amount and frequency of driving. Results of a Logliniar analysis on gender, age group, and change in driving frequency revealed significant effects. Exploring the Partial Associations revealed a significant two-way interaction between gender and change in driving frequency.
Thus the males are more likely to report reduced driving frequency. The two-way interaction between age group and change in driving frequency was likewise significant ($X^2(4) = 20.281, P<0.01$). There was no significant three-way interaction between age group, gender and change in driving frequency. Loglinear analysis on change in driving amount, gender and age group showed the same pattern. A significant two-way interaction of gender and change in driving amount ($X^2(2) = 9.092, P<0.05$) and two-way interaction between age group and change in driving frequency ($X^2(4) = 15.016, P<0.01$).

The participants who had reported reducing the amount and/or frequency of driving were presented with a list of possible reasons for having reduced ones driving. They were asked to rate how much each reason applied to them on a five point Likert type scale ranging from 1 = “not at all”, 2 = “poorly”, 3 “neither poorly nor good”, 4 = “good”, and 5 = “very good”. Mean scores were then calculated for females and males respectively for each reason. Hence, a mean score below approximately 2.8 indicates that the reason is rated as not relevant by the group, a mean score between approximately 2.8 and 3.2 is neutral, and a mean score above approximately 3.2 indicates that the reason is rated as relevant by the group.

Looking at the possible reasons for reducing ones driving the only reason which on average received a positive rating was “I have fewer activities I drive to” both for males (Mean = 3.57) and females (Mean = 3.48).

The participants were presented with a list of 17 driving situations and asked to indicate whether they, when possible, avoided each of these situations. In the analysis a total avoidance score was calculated which was the sum of all the situations avoided. Hence, the score ranges from 0 – 17 and the higher the score the more situations the person is avoiding.

![Figure 4 Avoidance score for females and males in the three age groups](image-url)
As can be seen in figure 4, females in all age groups have a higher avoidance score than males. In addition, it appears that the avoidance score is higher in the higher age groups for both genders. A Two-way independent ANOVA showed a significant main effect of gender ($F(2,737) = 15.218, P<0.01$) and a significant main effect of age group ($F(1,737)=41.011, P<0.01$). Helmert contrast results show that there is a significant difference between all three groups (Contrast estimate group 1 & 2 = -1.093, $P<0.01$, and contrast estimate group 2 & 3= -1.177, $P<0.01$). Thus the above trends are significant.

Participants who reported avoiding the following situations: Motorway, times and places with many bikes, left turns, driving when dark, and driving when feeling unwell, were presented with a list of possible reasons for avoiding the situation and asked to indicate to which extent each reason applied to them, using the same five point Likert type scale as when rating the reasons for reduced driving describes above.

As can be seen from figure 5 only 3.4% of female and male participants respectively avoided left turns, which is a total of 30 participants. Therefore no further analyses were conducted on this variable. The few participants reporting that they did not know if they avoided a situation were excluded so the percentages shown in figure 5 are based on the participants stating clearly “yes” or “no”.

Of the females 111 (25.3 %) and 50 males (11.3%) reported avoiding driving on motorway when possible (see figure 5). Looking at the individual reasons for avoiding driving on
motorway the same four reasons received positive ratings from both females and males: “I do not like to drive on motorway” (Mean=F:3.98, M:3.84), “I have no reason to drive on motorway” (Mean= F:4.08, M:3.63), “I prefer to drive another and prettier route” (Mean= F:3.81, M:3.85), and “The speed is too high on motorways” (Mean=F:3.60, M:3.78).

As shown in figure 5, 98 (22.5%) of the females and 80 (18.4%) of the males reported avoiding driving times and places with many bikes. Looking at the individual reasons for avoiding driving times and places with many bikes the females rated two reasons positive “I do not like to drive when there are many bikes” (Mean=3.51) and “I have no reason to drive places where there are many bikes” (Mean=3.57). The males did not rate any reasons positively.

222 (50.8%) of the females and 127 (29.1%) of the males reported avoiding driving when it is dark (See figure 5). Looking at the individual reasons for avoiding driving when it is dark both females and males rated three positively “I do not like to drive when it is dark” (Mean=F:4.16, M:3.83), “I have no reason to drive when it is dark” (Mean=F:3.96, M:3.71), and “I might as well move my activities to when it is light” (Mean=F:3.98, M:3.94). The rating for the reason “It takes too much concentration to drive when it is dark” was bordering to positive for both gender (Mean=F:3.21, M:3.24). Finally, females also rated “I feel unsecure when driving when it is dark” positively (Mean=3.42).

The situation that was avoided by most of the five situations shown in figure 5 was driving when feeling unwell. Of the females 360 (86.3%) and of the males 306 (75.2%) reported avoiding driving in this situation. Both genders rated the same three reasons for avoiding driving when feeling unwell as positive “I do not like to drive when I am feeling unwell” (Mean=F:4.16, M:4.00), “I have no reason to drive when I feel unwell” (Mean=F:4.10, M:3.87), and “I feel unsecure when driving when I feel unwell” (Mean=F:3.45, M:3.38). In addition, the rating for the reason “It takes too much concentration to drive when I am unwell” was bordering to positive for both genders (Mean=F:3.21, M:3.21).

**DISCUSSION**

The participants in this study were active drivers, most of them driving several times a week if not every day. Looking at the yearly mileage the males reported driving more kilometres per year than females. This gender difference declines in the older age groups which match the pattern commonly reported in the literature.
When looking at changes in driving exposure, results revealed that more males than females report having reduced both the amount and frequency of driving compared to 15 years ago in
the two youngest age groups. This can probably be explained by the fact that males tend to
drive more and thus have more driving to reduce. Also, a larger proportion of the females
were widowed. These women may not have same freedom to reduce their driving as they do
not have a spouse who can drive.
When looking at avoidance of traffic situations the females in all age groups avoid more
situations than males. Also, the older age groups avoid more situations than the younger. This
fits well with the literature finding that males self-regulate less than females and studies that
have found that drivers self-regulate more as they age. Unlike the study by Rimmö &
Hakamies-Blomqvist (2002) the gender difference did not decline in the higher age groups if
anything it appears to increase in this sample. There was however no significant interactions
between gender and age group so this trend is likely to be random.

The participants were asked to rate how well each reason on a list of 22 possible reasons for
reducing ones driving applied to them. Results showed that for both genders only one reason
“I have fewer activities I drive to” received a positive rating. This fits well with the studies by
Charlton et al. (2006) and Raitanen et al. (2003) that found that the most common reasons for
reduced driving was change in life style, retirement and being able to reach everything by car
whereas health was mentioned by fewer participants. In addition, the majority of the
participants in the present study rated their health positive and reported few difficulties with
cognitive functions which may also cause the internal reasons such as decline in vision and
concentration problems to be less relevant.

The participants were likewise asked to rate the personal relevance of possible reasons for
avoiding driving on motorway, when it is dark, time and places with many bikes, and when
feeling unwell.
Two reasons for avoidance were rated as relevant in nearly all driving situations by both
females and males “I do not like to drive in this situation” and “I have no reason to drive in
this situation”. The exception being “driving times and places with many bikes” where the
males rated none of the reasons positively. If you do not like to drive in a certain situation and
do not need to, then most people would probably avoid doing so. Thus, these two reasons are
most likely linked. Turning to the reason “I do not like to drive in this situation” it can be both
internally and externally motivated. The person may get this feeling because he or she feels
other road users drive carelessly, however, this feeling could also stem from functional
decline making driving in certain situations more demanding. Yet another explanation could
be lack of confidence in own driving causing the driver not to like driving in certain
situations. As people may put different meanings into the same reason, and internal reasons
may interact with external reasons, it is not possible to conclude if this most common reason
is internally or externally motivated. These subtle differences could possible emerge if
explored using qualitative methods.
When rating reasons for avoiding driving on motorway both genders also rated “I prefer to drive another and prettier route” positively. This looks like an external reason - it is the surroundings causing this choice and not an internal state. However, it cannot be ruled out that if a person “does not like” to drive on motorway because of functional decline of some sort, another route than motorway will certainly appear attractive. So again there could be an interaction between internal and external reasons for avoiding driving on motorway. Finally, both female and male participants rated “The speed is too high on motorways” positively. Again it could be both internally and externally motivated. Is it the other drivers that drive too fast and reckless or is it because it is unpleasant to drive at these speeds because the persons reaction time is not as good as it used to be? Both reasons could well express themselves as the feeling of not liking to drive on the motorway.

When avoiding driving when dark both genders, in addition, rated “I might as well move my activities to when it is light” positively. This again at first hand appears as an externally motivated reason. However, again if it is unpleasant to drive when it is dark because of say visual problems and the person therefore gets the feeling of not liking to drive when it is dark, then one would move ones activities if possible. So yet again it is not clear if the motivation for avoiding driving when it is dark is internal or external. In addition, the females rated “I feel unsecure when driving when it is dark” positively. This feeling of insecurity appears like an internally motivated reason but it could stem from, for example, being scared of an assault and thus not caused by problems with own functioning. Having said this, the rating for “It takes too much concentration to drive when it is dark” bordered to positive for both genders. This to a larger degree indicates internal motivation such as having difficulties seeing when dark and thus needing to concentrate more. Again, this needs to be explored using qualitative methods to better understand the interactions between internal and external motivations.

The reason “I feel unsecure when driving when feeling unwell” was also rated positively and by both genders. As discussed above it is unclear whether it is an internally or externally motivated reason. Also, “it takes too much concentration to drive when feeling unwell” was rated bordering to positive by both genders. Given that feeling unwell is a shift in a person’s internal state it seems plausible that the reasons given, although from the outside the same as when avoiding other situations, in this case are internally motivated.

There did not appear to be any gender differences other than females tending to rate more reasons positively than males. It cannot be ruled out that qualitative methods would find gender differences in whether the various reasons are internally or externally motivated or interact differently.
In conclusion, the results indicate that the general reduction of driving is, in many cases, caused by practical reasons i.e. “having fewer activities one drives to”. This result fits well with the existing literature.

The results indicate that the feeling of “not liking” to drive in a situation leads to the situation being avoided if possible. Lifestyle changes such as retirement may increase the options to avoid driving in a wider range of situations. If this feeling of “not liking” to drive in a driving situation stems from for example functional decline or illness causing a significant decrease in driving skill, it may be sensible to avoid these situations and the driver may thereby increase her or his safety on the road. However, if this feeling stems from lack of confidence despite adequate driving skills, then avoidance of the traffic situation may be a negative thing as it will lead to decrease in exposure and thereby routine in driving in these situations and perhaps others, possible leading to further decline in confidence and at worst perhaps premature driving cessation. It is not possible to conclude from the results of this study when and why this feeling of “do not like to drive” in certain situations arise. But it appears to be a central feeling involved in the vicious circle leading to premature driving cessation described in the introduction.

Further studies, exploring the interactions between internal and external motivations behind the reasons given for avoiding driving in different traffic situations, are needed. Using longitudinal designs may prove advantageous as it would make it possible to explore if and how motivations for self-regulation change in the process from the beginning of self-regulation to driving cessation. In order to capture the complex relationship between internal and external motivations for self-regulation qualitative interviews may be the most suitable approach, as this will allow the older drivers to use their own words to explain the process. Such studies may in addition reveal if there are gender and age differences in these processes.

Charlton et al. (2006) found that drivers who were not the principal driver in the household were more likely to show “self-regulatory avoidance behaviour”. In addition, D’ambrosio et al. (2008) found that female drivers who live alone are more likely to self-regulate than females who have others in the household who can drive, but they did not find the same trend among the male drivers. Females are more likely not to be the principal driver, at least until widowhood, which may be part of the explanation why they are often found to self-regulate more. Further studies exploring how the transition into widowhood affect self-regulation and perhaps even reasons for self-regulation, may also provide important information on the process of self-regulation and driving cessation and further explore differences in this process between sub-groups of older drivers.
REFERENCES


APPENDIX

List of possible reasons for reducing driving
- “because I do not like to drive any longer”
- “because I do not feel confident in my own driving”
- “because I have moved house”
- “because family members or friends have moved closer to”
- “because petrol is expensive”
- “because I now have better access to public transportation”
- “because I have fewer activities I drive to”
- “because I no longer drive just for the sake of the drive”
- “because it pollutes to drive”
- “because the other road users have become more reckless”
- “because the traffic has become more complex”
- “because there are too many bikes”
- “because I am scared of having an accident”
- “because I feel unsecure about driving”
- “because I previously have been in an accident”
- “because I feel physical discomfort when driving”
- “because driving takes too much concentration”
- “because I have difficulties orienting because of stiffness of the neck”
- “because my vision has declined”
- “because my reaction time has declined”
- “because I have been recommended to drive less by my GP or other doctor”
- “because a member of my family/or other person thinks it is best if I drive less”

List of possible reasons for avoiding driving on the motorway
- “because I do not like to drive on motorway”
- “because I do not feel confident in my own driving, when I drive in motorway”
- “because I have no reason to drive on motorway”
- “because I prefer to drive another and prettier route”
- “because the other drivers drive recklessly on the motorway”
- “because the speed is so high on the motorway”
- “because I am scared of having an accident”
- “because I feel unsecure about driving on motorway”
- “because I previously have been in an accident”
- “because it takes too much concentration to drive on motorway”
- “because my vision has declined”
- “because my reaction time has declined”
- “because a member of my family/or other person thinks it is best if I avoid driving on motorway”

**List of possible reasons for avoiding driving times and places with many bikes**
- “because I do not like to drive when there are many bikes”
- “because I do not feel confident in my own driving, when I drive where there are many bikes”
- “because I have no reason to drive where there are many bikes”
- “because the cyclists drive recklessly”
- “because I am scared of having an accident”
- “because I feel unsecure about driving when there are many bikes”
- “because I previously have been in an accident”
- “because it takes too much concentration to drive when there are many bikes”
- “because I have difficulties orienting because of stiffness of the neck”
- “because my vision has declined”
- “because my reaction time has declined”
- “because a member of my family/or other person thinks it is best if I avoid driving when there are many bikes”

**List of possible reasons for avoiding driving when it is dark**
- “because I do not like to drive when it is dark”
- “because I do not feel confident in my own driving, when I drive when it is dark”
- “because I have no reason to drive when it is dark”
- “because I have good access to public transportation which I prefer to use when it is dark”
- “because I might as well move my activities to when it is light”
- “because the roads I typically drive on have poor lightning”
- “because there are so many cyclists who drive without lights on”
- “because I am scared of having an accident”
- “because I feel unsecure about driving when it is dark”
- “because I previously have been in an accident”
- “because it takes too much concentration to drive when it is dark”
- “because my vision has declined”
- “because my reaction time has declined”
- “because a member of my family/or other person thinks it is best if I avoid driving when it is dark”

**List of possible reasons for avoiding left turns**
- “because I do not like to do left turns”
- “because I do not feel confident in my own driving, when I do left turns”
- “because the other drivers drive recklessly”
- “because I am scared of having an accident”
- “because I feel unsecure about doing left turns”
- “because I previously have been in an accident”
- “because it takes too much concentration to do left turns”
- “because I have difficulties orienting because of stiffness of the neck”
- “because my vision has declined”
- “because my reaction time has declined”
- “because I have difficulties judging when it is safe to drive forward”
- “because a member of my family/or other person thinks it is best if I avoid doing left turns”

List of possible reasons for avoiding driving when feeling unwell
- “because I do not like to drive when feeling unwell”
- “because I do not feel confident in my own driving, when I am feeling unwell”
- “because I have no reason to drive when feeling unwell”
- “because I am scared of having an accident”
- “because I feel unsecure about driving when feeling unwell”
- “because I previously have been in an accident”
- “because it takes too much concentration to drive when feeling unwell”
- “because it is difficult to keep the overview in traffic when feeling unwell”
- “because a member of my family/or other person thinks it is best if I avoid driving when feeling unwell”
QUANTIFYING THE INFLUENCE OF SOCIAL CHARACTERISTICS ON ACCIDENT AND INJURIES RISK: A COMPARATIVE STUDY BETWEEN MOTORCYCLISTS AND CAR DRIVERS

Allan Lyckegaard, DTU Transport, Kgs. Lyngby, Denmark
ally@transport.dtu.dk

Morten N. Olesen, Metroselskabet I/S, København S, Denmark
mnolesen@gmail.com

Tove Hels
ths@transport.dtu.dk

ABSTRACT

In the recent years many European countries have experienced an increase in the number of fatal traffic accidents with motorcycles. Bos et al. (2008) reports an increase from 17.4% to 21.1% of the total number of fatalities on powered two-wheelers in the European traffic. Several reasons for this have been suggested, among the most common is the hypothesis that during the last decade or so, the typical motorcyclist has become older, and as a result of the increase in age, the loss of physical ability in driving and orientation has resulted in the increase in the number of accidents (Værø 2008, SafetyNet 2009). In Denmark in the period 2002 to 2007, the average age of motorcycle owners increased from 42.3 to 45.2 years. In the same period, the average age of injured motorcyclists increased from 35.9 to 38.4 years, meaning that the average injured motorcyclist has become younger in this period.

In this analysis we establish relationships between social and demographic characteristics and the probability of being in an accident and being injured in an accident. Logistic regression was applied to both motorcyclists and car drivers with the purpose of calculating the odds ratio with the car drivers as the control group. The available data for the regression consisted of accident and injury data for motorcyclists and car owners in the period of interest as well
the social and demographic parameters: age, gender, income, educational level and family status. The odds ratio calculations showed that the risk of being in an accident or in an injury accident decreased with age, educational level, and income. Furthermore, the risk of being in an accident was 1.72 to 1.96 times higher and the risk of being in an injury accident was 1.38 to 1.44 times higher for men compared to women. For motorcyclists compared to car drivers, the risk of being in an accident was 1.44 to 1.78 times higher and the risk of being in an injury accident was 2.29 to 3.16 times higher. Singles showed an increased risk of 1.25 to 1.87 times higher for being in an accident and 1.50 to 2.25 times higher risk for being in an injury accident when comparing to person a couple with children.

**INTRODUCTION**

In Denmark, as well as in the rest of Europe, the period 2002 to 2007 had a high economical growth enabling more people to buy a motorcycle. The common observation expressed in the media was that these motorcycles were bought as a leisure time toy and not with purpose of transportation though there is no evidence that this should be the case. An analysis of the net growth between 2002 and 2007 showed that the largest growth happened in the group of people with an income in the highest 40% percentile. A further analysis of the age distributions revealed a shift in these distributions towards older owners. Figure 1 and 2 shows the age distribution of motorcycle owners and owners of both motorcycle and car, respectively.

Figure 1. Age distribution of owners of motorcycles in 2002 and 2007.
The analysis of the income and age distributions suggests that there has been a change in the social and demographic characteristics of the motorcycle owner in Denmark from 2002 to 2007. In the same period of time, the number of killed and seriously injured motorcyclists has increased. The fact that more motorcyclists are getting killed and seriously injured is opposite the trend seen in Danish traffic in total. Since the mid 1970’s, there has been a downward trend in the number of killed and seriously injured persons in traffic in Denmark, but the trend for the group of motorcyclists is clearly opposite in the period from 2002 to 2007. Up until 2005 the number of killed motorcyclists has been stable around 5-6% but in 2006 the number was more than 7% and in 2007 the number was almost 9%. These numbers should be compared to the estimated 2% of the traffic that motorcycles and mopeds represent (ETSC 2007). In the rest of Europe, the same pattern with more people buying a motorcycle and more people getting in accidents with motorcycles can be seen. From 2000 to 2005 the number of fatalities on powered two-wheelers in the EU-15 countries went from 3739 to 4047 corresponding to 17.4 % and 21.1 % of the total number of people killed in traffic (Bos et al. 2008).

In Denmark, an analysis of the numbers given in DRTAIB (2009), shows that the age of the motorcyclists increased from 42.3 to 45.2 years in the period 2002 to 2007 (N_{2002}=19548 and N_{2007}=38266). In the same period, the average age of injured motorcyclists increased from 35.9 to 38.4 years (N_{2002}=222 and N_{2007}=272), meaning that the average injured motorcyclist has become younger when comparing 2007 to 2002. Figure 3 shows the age distribution of two three-year periods, 2002-2004 and 2005-2007, for killed and seriously injured motorcyclists. The distribution shift towards the older segments when comparing the period 2002-2004 with 2005-2007.
Figure 3. Age distribution of killed and seriously injured motorcyclists in the two periods 2002-2004 and 2005-2007.

The Danish Road Traffic Accident Investigation Board made an in-depth analysis of 41 motorcycle accidents (Værø 2008). The analysis showed that the main causes of accidents were speeding and willingness to take risks, i.e. factors which relates to the behaviour of the motorcyclist. As only 41 accidents were analysed and they only cover the year 2008, it is not possible to draw strong statistical conclusions or say something about development over time.

This paper aims at building on top of the in-depth analysis of the 41 accidents and the simple analysis of motorcycle accident data performed by the Danish Road Traffic Accident Investigation Board (Værø 2008, DRTAIB 2009) by focusing on a quantitative analysis of a large number of accidents with data on individual level. The analysis is an epidemiologic analysis of which factors contribute to increased risk based on socio-demographic data on all motorcyclists and car drivers in Denmark in the period 2002 to 2007.

**METHOD**

The overall goal with the analysis is to compute the odds ratio for being in an accident or in an accident with personal injury (fatalities are considered as injuries in this study) based on a logistic regression of a number of socio-demographic factors. The design of the study was made as a case/control study with three groups.

**Sample**

The data used in the analysis comes from Statistics Denmark’s collection of register data and are linked to individuals such that various registers of interest can be merged. For this
analysis, all owners of a car or a motorcycle were extracted from the central vehicle register and split into three subgroups: Car, motorcycle (MC) or car/motorcycle owner, see table 1.

Table 1. Number of observations in the individual groups used for the analysis.

<table>
<thead>
<tr>
<th>Ownership</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car</td>
<td>1,076,127</td>
<td>1,566,716</td>
<td>1,577,078</td>
<td>1,602,615</td>
<td>1,630,135</td>
<td>1,652,454</td>
</tr>
<tr>
<td>MC</td>
<td>19,548</td>
<td>27,446</td>
<td>29,032</td>
<td>31,070</td>
<td>34,449</td>
<td>38,266</td>
</tr>
<tr>
<td>Car &amp; MC</td>
<td>31,698</td>
<td>48,219</td>
<td>52,536</td>
<td>59,204</td>
<td>67,198</td>
<td>75,807</td>
</tr>
</tbody>
</table>

This study is designed as a case/control study. The car owners are included in the analysis as a control group since this group of road users share the same space on the road as motorcycles and are subject to almost the same type of legislation. The sample is almost the complete population of motorcycle and car owners in Denmark except for 2002 where almost 500,000 vehicles were not registered correctly and therefore missing in the data. Chi-square tests on gender, income and family type showed that there was a small but significant difference (95%) from year to year for all years. For education, a T-test showed a significant difference (99%) in the length of education when comparing 2002 and 2007. Due to the large number of observations even small differences of up to 0.1 year in education length was significant in the T-test. Based on these tests it was decided that the observations from 2002 should be included in the analysis as 2002 was no more different from the other years than any other year.

For each of the individuals included in one of the three groups (Car, MC, Car & MC), gender (male/female), age (in years), income (in DKK before taxation), education length (in years), family relationship (in a couple with children/in a couple without children/single with children/single without children) and presence in the police accidents register is extracted. In the police accident register, accidents with material damage, personal injury and fatal accidents are registered making it possible to distinguish between any type of accident and accidents involving personal injury. It is known that the accident data is suffering from underreporting (Elvik and Mysen 1999, Amoros et al. 2006), but as almost all severe accidents are recorded in the police accident register and there is no systematic difference in the underreporting between case and control groups, the data quality is good enough to use in the study of risk. Furthermore, persons involved in an accident as a passenger on either a motorcycle or in a car are left out as they have no or very little influence on how the car or motorcycle is controlled so it is not relevant to include them in the analysis of risk.
Data analysis

The basis of the risk analysis is the logistic regression which is known from many epidemiologic studies. The purpose is to determine a person’s probability $p_i$ of being in an accident or an accident with personal injury given a number of socio-demographic factors $\beta_{\text{vehicle}}$, $\beta_{\text{gender}}$, $\beta_{\text{age}}$, $\beta_{\text{education}}$, $\beta_{\text{income}}$ and $\beta_{\text{family}}$ for each individual. The relationship between the probability $p_i$ and the socio-demographic factors $\beta$ is modelled by logistic regression (Madsen and Thyregod 2010):

$$\log \left( \frac{p_i}{1-p_i} \right) = \beta_0 + \beta_{\text{vehicle}} x_{\text{vehicle}} + \beta_{\text{gender}} x_{\text{gender}} + \beta_{\text{age}} x_{\text{age}} + \cdots + \beta_{\text{family}} x_{\text{family}}$$

With this model it is assumed that the underlying unobserved behaviour is equally distributed over both the categorical and the continuous variables.

The odds ratio is an approximation to relative risk when the event under investigation is a rare event. As accidents in general are rare, the assumption that the odds ratio approximates the relative risk is valid. When the odds ratio is estimated from (1), it is possible to separate out how each factor influence the odds ratio with respect to a reference group, as the individual contribution of the $\beta$-factors is linear and without interactions between the factors. The chosen reference group will be incorporated into the intercept $\beta_0$.

The data analysis and model building was performed in SAS 9.1 using PROC LOGISTIC to compute the parameters of the logistic regression and the corresponding odds ratios.

RESULTS

The results of the logistic regression for both data on all accidents and data on accidents with personal injury are given in this section.

Odds ratio for all accidents

In table 2, the parameter estimates of the logistic regression over all accidents are given. As the data contains categorical variables it is necessary to define a reference group. The reference group (female car owners in a couple with children) is incorporated into the intercept and forms the baseline which the other groups are compared to. The estimated parameters are therefore negative when a group has lower probability compared to the reference group and positive parameter estimates when a group has higher probability compared to the reference group.
Table 2. Parameter estimates for the logistic regression for all accidents. Parameter estimates marked with * is estimated to level of significance lower than 95%. Cells marked with – could not be estimated.

<table>
<thead>
<tr>
<th></th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.95</td>
<td>-3.91</td>
<td>-4.04</td>
<td>-4.20</td>
<td>-4.09</td>
<td>-3.71</td>
</tr>
<tr>
<td>MC</td>
<td>0.19</td>
<td>0.24</td>
<td>0.12*</td>
<td>0.16*</td>
<td>0.12*</td>
<td>0.24</td>
</tr>
<tr>
<td>Car&amp;MC</td>
<td>0.10*</td>
<td>0.09*</td>
<td>0.13</td>
<td>0.17</td>
<td>0.17</td>
<td>0.04*</td>
</tr>
<tr>
<td>Gender (male)</td>
<td>0.32</td>
<td>0.33</td>
<td>0.28</td>
<td>0.27</td>
<td>0.34</td>
<td>0.33</td>
</tr>
<tr>
<td>Age</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.03</td>
<td>0.03</td>
<td>-0.03</td>
<td>-0.03</td>
</tr>
<tr>
<td>Single w.o. children</td>
<td>0.21</td>
<td>0.11</td>
<td>0.16</td>
<td>0.17</td>
<td>0.15</td>
<td>0.07</td>
</tr>
<tr>
<td>Single w. children</td>
<td>0.16</td>
<td>0.36</td>
<td>0.31</td>
<td>0.28</td>
<td>0.22</td>
<td>0.27</td>
</tr>
<tr>
<td>In a couple w.o. children</td>
<td>-0.19</td>
<td>-0.21</td>
<td>-0.24</td>
<td>-0.19</td>
<td>-0.17</td>
<td>-0.20</td>
</tr>
<tr>
<td>Income, 10^-7</td>
<td>-4.57</td>
<td>-2.49</td>
<td>-3.27</td>
<td>-</td>
<td>-2.65</td>
<td>-</td>
</tr>
<tr>
<td>Education</td>
<td>-0.05</td>
<td>-0.05</td>
<td>-0.05</td>
<td>-0.06</td>
<td>-0.07</td>
<td>-0.09</td>
</tr>
<tr>
<td>R²</td>
<td>0.0008</td>
<td>0.0007</td>
<td>0.0006</td>
<td>0.0005</td>
<td>0.0006</td>
<td>0.0006</td>
</tr>
</tbody>
</table>

From table 2 it can be seen that the accident probability for both MC and Car&MC is higher than the reference group Car as the parameters estimates are greater than zero. Not all the estimates are significant to a level of 95% significance but the overall trend is that groups MC and Car&MC have a higher probability compared to the group Car. Furthermore, the factors for males and singles increase the probability of being in an accident and the factors age, in a couple with children, income and education decrease the probability of being in an accident. These results do not contradict what has been seen before.

The R² is very small so the model only explains very little of a person’s accident probability which is what could be expected as the socio-demographic factors only to some degree explain something about a person’s behaviour in traffic or level of aggression. The goal is to find a relationship between risk and demographic factors, not building an accident model, so the small R² only show that there are more factors which also influence this risk.

The odds ratios for all accidents are given in table 3. For the categorical variables, the groups are compared to the reference group, and for the continuous variables age, education and income, the slope of the linear relationship is evaluated by comparing two points on the line, e.g. Age and Age+1. The conclusion is similar to the one based on table 2, namely, that the groups MC and Car&Mc have a higher risk of being in an accident. Males have almost twice the risk compared to women and singles have a higher risk than persons in a couple. Higher
age and longer education is related to a lower risk of being in an accident, whereas higher income does not seem to play any role.

Table 3. Odds ratio estimates for all accidents based on logistic regression. Cells marked with – could not be estimated.

<table>
<thead>
<tr>
<th></th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC vs. Car</td>
<td>1.61</td>
<td>1.78</td>
<td>1.44</td>
<td>1.64</td>
<td>1.52</td>
<td>1.68</td>
</tr>
<tr>
<td>Car&amp;MC vs. Car</td>
<td>1.47</td>
<td>1.53</td>
<td>1.45</td>
<td>1.66</td>
<td>1.59</td>
<td>1.37</td>
</tr>
<tr>
<td>Male vs. Female</td>
<td>1.89</td>
<td>1.92</td>
<td>1.75</td>
<td>1.72</td>
<td>1.96</td>
<td>1.93</td>
</tr>
<tr>
<td>Age vs. Age+1</td>
<td>0.98</td>
<td>0.97</td>
<td>0.97</td>
<td>0.98</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>Single w.o. children vs. In a couple w. children</td>
<td>1.47</td>
<td>1.46</td>
<td>1.49</td>
<td>1.53</td>
<td>1.42</td>
<td>1.25</td>
</tr>
<tr>
<td>Single w. children vs. In a couple w. children</td>
<td>1.40</td>
<td>1.87</td>
<td>1.73</td>
<td>1.71</td>
<td>1.52</td>
<td>1.52</td>
</tr>
<tr>
<td>In a couple w.o. children vs. In a couple w. children</td>
<td>0.99</td>
<td>1.06</td>
<td>1.00</td>
<td>1.07</td>
<td>1.03</td>
<td>0.95</td>
</tr>
<tr>
<td>Income vs. income+1000</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
</tr>
<tr>
<td>Education vs. education+1</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.94</td>
<td>-</td>
<td>0.91</td>
</tr>
</tbody>
</table>

The odds ratio calculations show that the groups MC and Car&MC, i.e. individuals containing at least one motorcycle, have a higher risk of being in an accident than the control group Car. Likewise, young persons and males have a higher risk compared to older and women which is consistent with what is reported in Brems and Munch (2008). Furthermore, singles show to have higher risk and the more education a person have, the lower the risk.

Odds ratio for accidents with personal injury

In table 4, the parameters estimates of the logistic regression over accidents with personal injury are given. As in the analysis above, the reference group incorporated into the intercept was female car owners in a couple with children.
Table 4. Parameter estimates for the logistic regression for accidents with personal injury. Parameter estimates marked with * is estimated to level of significance lower than 95%. Cells marked with – could not be estimated.

<table>
<thead>
<tr>
<th></th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-4,14</td>
<td>-4,10</td>
<td>-4,11</td>
<td>-4,20</td>
<td>-4,25</td>
<td>-3,92</td>
</tr>
<tr>
<td>MC</td>
<td>0,38</td>
<td>0,44</td>
<td>0,32</td>
<td>0,35</td>
<td>0,43</td>
<td>0,47</td>
</tr>
<tr>
<td>Car&amp;MC</td>
<td>0,29</td>
<td>0,28</td>
<td>0,20</td>
<td>0,31</td>
<td>0,25</td>
<td>0,18</td>
</tr>
<tr>
<td>Gender (male)</td>
<td>0,16</td>
<td>0,18</td>
<td>0,18</td>
<td>0,16</td>
<td>0,16</td>
<td>0,17</td>
</tr>
<tr>
<td>Age</td>
<td>-0,03</td>
<td>-0,04</td>
<td>-0,04</td>
<td>-0,04</td>
<td>-0,03</td>
<td>-0,04</td>
</tr>
<tr>
<td>Single w.o. children</td>
<td>0,25</td>
<td>0,22</td>
<td>0,20</td>
<td>0,26</td>
<td>0,20</td>
<td>0,19</td>
</tr>
<tr>
<td>Single w. children</td>
<td>0,20</td>
<td>0,41</td>
<td>0,33</td>
<td>0,32</td>
<td>0,19*</td>
<td>0,33</td>
</tr>
<tr>
<td>In a couple w.o. children</td>
<td>-0,19</td>
<td>-0,23</td>
<td>-0,26</td>
<td>-0,21</td>
<td>-0,17</td>
<td>-0,28</td>
</tr>
<tr>
<td>Income, $10^6$</td>
<td>-1,04</td>
<td>-0,52</td>
<td>-0,62</td>
<td>-0,62</td>
<td>-0,35</td>
<td>-</td>
</tr>
<tr>
<td>Education</td>
<td>-0,05</td>
<td>-0,05</td>
<td>-0,07</td>
<td>-0,07</td>
<td>-0,07</td>
<td>-0,10</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0,0004</td>
<td>0,0005</td>
<td>0,0004</td>
<td>0,0004</td>
<td>0,0003</td>
<td>0,0004</td>
</tr>
</tbody>
</table>

Again it is clear that the groups MC and Car&MC have a higher probability of being in an accident with personal injury. Similarly, males and singles have a higher probability compared to the reference group and age, income and education length decrease the probability of being in an accident with personal injury.

Table 5. Odds ratio estimates for accidents with personal injury. Cells marked with – could not be estimated.

<table>
<thead>
<tr>
<th></th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC vs. Car</td>
<td>2,86</td>
<td>3,16</td>
<td>2,29</td>
<td>2,75</td>
<td>3,05</td>
<td>3,09</td>
</tr>
<tr>
<td>Car&amp;MC vs. Car</td>
<td>2,60</td>
<td>2,71</td>
<td>2,04</td>
<td>2,65</td>
<td>2,56</td>
<td>2,30</td>
</tr>
<tr>
<td>Male vs. Female</td>
<td>1,39</td>
<td>1,44</td>
<td>1,43</td>
<td>1,38</td>
<td>1,37</td>
<td>1,41</td>
</tr>
<tr>
<td>Age vs. Age+1</td>
<td>0,97</td>
<td>0,97</td>
<td>0,96</td>
<td>0,97</td>
<td>0,97</td>
<td>0,96</td>
</tr>
<tr>
<td>Single w.o. children vs. In a couple w. children</td>
<td>1,67</td>
<td>1,86</td>
<td>1,61</td>
<td>1,86</td>
<td>1,53</td>
<td>1,53</td>
</tr>
<tr>
<td>Single w. children vs. In a couple w. children</td>
<td>1,59</td>
<td>2,25</td>
<td>1,82</td>
<td>1,98</td>
<td>1,50</td>
<td>1,76</td>
</tr>
<tr>
<td>In a couple w.o. children vs. In a couple w. children</td>
<td>1,08</td>
<td>1,19</td>
<td>1,01</td>
<td>1,17</td>
<td>1,05</td>
<td>0,96</td>
</tr>
<tr>
<td>Income vs. income+1000</td>
<td>1,00</td>
<td>1,00</td>
<td>1,00</td>
<td>1,00</td>
<td>1,00</td>
<td>-</td>
</tr>
<tr>
<td>Education vs. education+1</td>
<td>0,95</td>
<td>0,95</td>
<td>0,94</td>
<td>0,94</td>
<td>0,93</td>
<td>0,90</td>
</tr>
</tbody>
</table>
Table 5 shows the odds ratio for being in an accident with personal injury compared to the reference group. The pattern in the result is the same as for all accidents, i.e. the groups MC and Car&MC have a higher risk than the group Car. Singles and males again show higher risk, but where a person in a couple without children had a lower risk before, such a person now have a higher risk compared to the reference group. Length of education and age again decrease the risk and income still does not play any role.

Compared to the results for all accidents, it is worth noticing that the risk for motorcyclists being in an accident with personal injury is higher compared to motorcyclist being in any kind of accident. This indicates that when a motorcyclist in involved in an accident, then there is a higher risk for the motorcyclist to be injured compared to a person driving in a car. Similar results have been found in SWOV (2009).

**DISCUSSION AND CONCLUSION**

It is evident that the risk of motorcyclists being involved in an accident and getting injured in an accident is higher compared to car drivers. Furthermore, it was observed that there in the period 2002 to 2007 were no change in the relation between age and risk of being involved in an accident or the risk of being in an accident with personal injury. This means, that the increase in the number of accidents with motorcyclists in the period 2002 to 2007 could not be ascribed to older motorcyclists. In the whole period from 2002 to 2007 the motorcyclist with the highest risk of being involved in an accident and in an accident with personal injury is a young, single male without children, low income and only a short education.

Even though the models constructed above were significant, it should be noted that they only explain a very small amount of the variance seen in the data, i.e. $R^2$ is small. This suggests that the regression models could be extended further to give a more precise picture of the risk of being involved in an accident. More socio-demographic factors could be included such as the criminal records of the individuals or the medical history. The criminal records could indicate if a person perhaps is more likely to break the law in general and therefore maybe also break the traffic law. The medical history could give access to the list of medical drugs that a person may be under influence of or the person’s health, in particular if the person suffer from an illnesses which may influence the behaviour in traffic. Also the drivers experience would be of interest and the type and motor size of the bike.

In the models presented in this paper, only linear expressions without interactions have been used. As there is a large data material, it may be possible to include interactions in the models since there may exist correlations between some of the factors included.
One important issue that has not been discussed in this paper is exposure in terms of the yearly travelled distance. It is assumed that all individual owning a car or a motorcycle have the same exposure which is not the case. Unfortunately, yearly access to this kind of data is not easy on an individual level, although some data exist from vehicle inspections. To be able to develop a more precise model of the data, estimates for the yearly travelled distance on an individual level is needed.

REFERENCES


Katerina Chrysostomou
Centre for Research and Technology Hellas
Hellenic Institute of Transport
chrysostomou@certh.gr

QUANTIFYING DRIVER DISTRACTION –
THE CASE STUDY OF THESSALONIKI’S RING ROAD
Abstract

Driving is a task that involves concurrent execution of multiple tasks and so, intense attention. At the same time, distractions are everywhere and are potentially responsible for causing traffic accidents. Approximately 25-30% of accidents are caused by driver distraction and while numbers are important, they may not indicate the actual size of the problem, since the recognition of the distraction and its role in the crash can very difficultly be detected.

Driver distraction includes lifestyle issues, not just driving issues, including the almost natural tendency of observing objects, events or activities that are new, unusual or attractive. It appears due to a wide range of events, objects, such as advertising billboards, and activities outside and inside the vehicle. While driving, the attention is constantly divided into a multitude of activities and events, reducing the ability of the drivers to process information and thus lowering the threshold of road safety.

In this paper a methodological approach is proposed, in order to quantify driver distraction across the road environment. Specifically, driver distraction has been monitored during an experimental process in Thessaloniki’s ring road, using an on-vehicle eye tracking system. The main idea has been to identify the elements that attract the gaze of the driver away from the driving task, across an urban motorway and to assess the frequency and duration of this phenomenon. For this purpose gaze tracking cameras and a scene camera capturing the vision field of the driver have been used and the data has been fused using adequate software. The detailed description of the methodology, the statistical analysis of the results and the conclusions drawn, related to driver distraction, will be presented.

Key words: driver distraction, road safety, advertising billboards

Research domain: Transport safety
1. INTRODUCTION

1.1 BACKGROUND INFORMATION

The price people pay for the sake of mobility has always been very high. Since 1970, over 1.65 million of Europeans have lost their lives in road accidents and only in 2009, more than 35,000 people lost their lives in road accidents on the streets of the European Union, i.e. the equivalent of a medium sized city, and at least 1,500,000 people were injured (EU Energy and Transport in Figures, 2010). The cost to society is enormous, reaching about 130 billion Euros in 2009\(^1\).

It is important that the most affected age range is between 16 - 24 years in which traffic accidents are the leading cause of death. Approximately 70,500 persons aged 16-24 were killed in road traffic accidents, in 14 EU countries between 1996 - 2005, almost quarter of all road fatalities in those countries. Young people (aged 16-24 years) are at almost twice the risk of being killed in a road accident than the average member of the population across the EU-18 countries as a whole. (CARE - EU road accidents database).

The problem of road safety is multiparametric and thus it is quite difficult to determine the exact causes of road accidents. Accidents occur undoubtedly due to a combination of limitations of human performance, undesirable driver behavior, road conditions and other circumstances.

There is a generally accepted classification of the key factors impacting on road safety. Contributing factors are traditionally categorized in the following three groups:

- Factors affecting the drivers (road users)
- Factors affecting the vehicles and
- Factors affecting the road environment (including weather, light and traffic conditions).

These road safety factors are not completely independent of each other. Road accidents are closely connected to all three of them, varying in the weight of each factor in each circumstance. In most cases two or all three of these factors contribute to an accident. Human factor, alone or in combination with the other two factors, is the main cause of road accidents.

\(^1\) Based on the statistical value of life as calculated in the HEATCO study (6\(^{th}\) Framework Programme for Research and Technological Development).
1.2 OVERVIEW OF DRIVER DISTRACTION

Driver distraction is associated with all three traditional factors affecting road safety; the road environment, the vehicle and the driver, as distraction may be caused by an element on the road, a device in the vehicle or the driver himself with a thought or act. In order to assess driver distraction it is needed to investigate and understand all interactions between these three factors.

Driver distraction is gaining more and more attention from the media, governments, industry, security agencies and the public. The initial concerns arose from the rapid development and the highly visible use of mobile phones while driving. But there are many other sources of distraction deriving both from the interior and the exterior environment.

Definition of Driver Distraction

There is a number of attempts listed to define driver distraction. In the 1st International Conference on Distracted Driving (2005), the scientific community agreed on the following definition for driver distraction: “Distraction involves a diversion of attention from driving, because the driver is temporarily focusing on an object, person, task, or event not related to driving, which reduces the driver’s awareness, decision-making, and/or performance, leading to an increased risk of corrective actions, near-crashes, or crashes.”

More succinctly, driver distraction could be defined as any activity that turns driver’s attention away from the driving task. Accordingly, any distraction, from adjusting the radio to making a phone call, could possibly contribute to an accident since it diverts driver’s attention from the primary task of operating the vehicle.

Driving is a complex task that requires intense attention. Apart from controlling the vehicle, which is a complex task in itself, including actions such as steering, acceleration, braking, speeding, lane adjustment, driving maneuvers, the driver has to determine why, when and where he wants to drive. He has to navigate to the destination, make sure he takes the correct turns, explores the risks, avoids other vehicles and keeps the appropriate distance to lead vehicles. Furthermore, many secondary tasks such as tuning the radio, eating / drinking, grooming, reading signs, talking with a passenger (-s), smoking, etc, are performed while driving. This requires constant attention, goal management and use of memory.

According to Michon, driving consists of three hierarchically ordered levels (Michon, 1971):

- strategic
- tactical and
- operational.
On the strategic (navigation) level, the goals of the trip are set and the route and departure time are determined. On the tactical (guidance) level, the driver has to follow the road, maintain a steady speed and keep enough distance to other vehicle. On operational (control) level, the driver controls the vehicle by pressing the gas pedal and the brakes, turning the steering wheel and using the vehicle controls. These levels can be active at the same time and can influence each other.

Concurrent execution of multiple tasks would not be a matter if drivers could deal with them simultaneously - that is, if people could process simultaneously information coming from different sources. However, people are serial information processors and may participate in only one thing at a time. While most people think they can do many things simultaneously, what is commonly called multi-tasking actually involves a rapid shift of attention from one cognitive task to another (Smiley 2005).

Driving is already a task requiring intense attention to do many things simultaneously. Consequently, when another action or event distracts the driver and focus on the primary tasks of driving is impaired. Obviously, depending on factors such as the duration of distraction, traffic prevailing conditions, relative positioning of the vehicle in traffic, road geometry and driver experience and status, the consequences can be devastating.

Types of Driver Distraction

In general, there are 4 major types of distraction and are similar to the kinds of distractors that cause them. The types of distraction that can distract drivers are (Azman A., Meng Q. 2009):

- Visual distraction that happens when the driver's eyes are off the road, for example if the driver is looking at a billboard or a pedestrian by the roadside.
- Cognitive distraction that occurs when for example the driver is thinking of something not related to driving task and safety (e.g what to cook for dinner).
- Acoustic distraction which occurs when an audio source distracts drivers, impairing their ability to hear sounds related to driving.
- Manual distraction, that happens when the hands of the driver are not on the steering wheel, e.g. when changing the CD on audio system of the car.

Frequency of Driver Distraction

The U.S. National Highway Traffic Safety Administration (NHTSA) conducts extensive research on driver distraction and estimates - conservatively - that driver distraction contributes to 25-30% of accidents (Stutts, J. 2005).
According to an official announcement of the Traffic Division of Greek Police about accident data concerning the first 9 months of 2010, driver distraction is the main cause of 1,502 accidents out of 10,183 accidents that were caused due to drivers. According to the source, distraction is factor number one, followed by violation of priority (1,319), drunk driving (1,148) and speeding (917).

The figures following are an analysis of traffic accidents as they were recorded by Traffic Police (Table 1). It is noteworthy that estimates based on police reports are hampered by the fact that most jurisdictions do not even include driver distraction as a category for police to check when assigning a potential cause (Sundeen 2005). That’s why while examining these data, the difficulty of identifying driver distraction in an accident must be taken into account. It is difficult to assign the cause of an accident to driver distraction since the cause is not self-evident and the accusation is based on the expert’s conjecture. Moreover, many of the causes listed below may be a consequence of driver distraction. For example, sudden braking may be justified due to an event that led to distraction.

According to the data, driver distraction is a major cause of traffic accidents in Greece with approximately 12%. The rates are ever higher if other causes such as traffic signs violation, the violation of priority and other reasons obviously related to driver distraction are also taken into account. As a result driver distraction seems to be a major cause of traffic accidents in Greece referring up to 50% of the causes.

<table>
<thead>
<tr>
<th>Causes</th>
<th>2009</th>
<th>%</th>
<th>2010</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speeding</td>
<td>1,382</td>
<td>8,45</td>
<td>582</td>
<td>6,92</td>
</tr>
<tr>
<td>Improper overtaking</td>
<td>281</td>
<td>1,72</td>
<td>191</td>
<td>2,27</td>
</tr>
<tr>
<td>Moving in the opposite way</td>
<td>999</td>
<td>6,11</td>
<td>542</td>
<td>6,44</td>
</tr>
<tr>
<td>Violation of priority</td>
<td>1448</td>
<td>8,85</td>
<td>861</td>
<td>10,23</td>
</tr>
<tr>
<td>No safe distances</td>
<td>340</td>
<td>2,08</td>
<td>258</td>
<td>3,07</td>
</tr>
<tr>
<td>Driver distraction</td>
<td>2243</td>
<td>13,71</td>
<td>965</td>
<td>11,47</td>
</tr>
<tr>
<td>Others causes due to drivers</td>
<td>5682</td>
<td>34,73</td>
<td>3195</td>
<td>37,96</td>
</tr>
<tr>
<td>Causes due to passengers</td>
<td>137</td>
<td>0,84</td>
<td>90</td>
<td>1,07</td>
</tr>
<tr>
<td>Causes due to the vehicle</td>
<td>460</td>
<td>2,81</td>
<td>207</td>
<td>2,46</td>
</tr>
<tr>
<td>Causes due to the road or the weather</td>
<td>1494</td>
<td>9,13</td>
<td>661</td>
<td>7,85</td>
</tr>
<tr>
<td>Causes due to pedestrians</td>
<td>1893</td>
<td>11,57</td>
<td>864</td>
<td>10,27</td>
</tr>
<tr>
<td>Total</td>
<td>16,359</td>
<td>100</td>
<td>8416</td>
<td>100</td>
</tr>
</tbody>
</table>

Source: Greek Police (www.astynomia.gr)

Similarly, in fatal crashes (Table 2) driver distraction plays an important role giving an indication of the risk due to driver distraction.
Table 2: Fatal crashes analysis (1st Semester 2010)

<table>
<thead>
<tr>
<th>Causes</th>
<th>2010</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speeding</td>
<td>108</td>
<td>20,57</td>
</tr>
<tr>
<td>Improper overtaking</td>
<td>10</td>
<td>1,90</td>
</tr>
<tr>
<td>Moving in the opposite way</td>
<td>72</td>
<td>13,71</td>
</tr>
<tr>
<td>Violation of priority</td>
<td>60</td>
<td>11,43</td>
</tr>
<tr>
<td>Driver distraction</td>
<td>23</td>
<td>4,38</td>
</tr>
<tr>
<td>Traffic violation</td>
<td>3</td>
<td>0,57</td>
</tr>
<tr>
<td>Careless driving</td>
<td>110</td>
<td>20,95</td>
</tr>
<tr>
<td>Others causes due to drivers</td>
<td>74</td>
<td>14,10</td>
</tr>
<tr>
<td>Still investigated</td>
<td>22</td>
<td>4,19</td>
</tr>
<tr>
<td>Causes due to passengers</td>
<td>0</td>
<td>0,00</td>
</tr>
<tr>
<td>Causes due to pedestrians</td>
<td>32</td>
<td>6,10</td>
</tr>
<tr>
<td>Causes due to the vehicle</td>
<td>1</td>
<td>0,19</td>
</tr>
<tr>
<td>Causes due to the road or the weather</td>
<td>10</td>
<td>1,90</td>
</tr>
<tr>
<td>Total</td>
<td>525</td>
<td>100</td>
</tr>
</tbody>
</table>

*Source: Greek Police ([www.astynomia.gr](http://www.astynomia.gr))*

Sources of distraction - Factors affecting Driver Distraction

Two important distinctions appear in literature: the driver distraction due to distractions inside and outside of the vehicle. The internal sources of distraction refer to what is happening inside the car (including driver's own actions) covering everything from adjusting the radio / CD player, conversations with passengers, use of mobile phones (or other communication and technology systems), to daydreaming.

External sources include everything out of the car; from the weather to billboards, children playing in the street and so on. Even though most (but not all) external distractions tend to be visual, it can be argued that, for example, police or ambulance sirens can also distract drivers.

Driver distraction due to external sources outside is getting very important. The urban environment is increasingly complex and crowded with factors who can act as sources of distraction, such as buildings, antennas, plants, screens and ads on buses, taxis, billboards and bus stops. Moreover, ads are becoming more and more widespread and significant since advertisers trying hard to attract attention.

There are many parameters affecting driver distraction such as various demographic characteristics (age and gender), as well as road environment factors (Lightning, Weather Conditions, Road Type, Road Profile, Traffic volume, Pavement condition, Traffic control systems, Interchanges) and vehicle characteristics.
Methods of evaluating Driver Distraction

The critical question is whether these distractions increase crash risk. To evaluate driver distraction the cooperation of multiple professionals is required to properly interpret the results of investigations. Engineering issues mainly deal with the choice of methods used in the experimental design, the definition of objectives and targets of an investigation and the evaluation of the results about road safety by proposing solutions that could be adopted by the automotive industry. However it is an issue of human sciences (doctors and psychologists) to evaluate and define the ways to detect distraction and to identify related to distraction human actions.

The effects of driver distraction can be measured through various research methods, the most popular of which fall into the following categories:

- **Crash-based studies** for example Correlation studies or accident analyses, that begin with the outcome (the collision itself) and endeavor to reconstruct what factors were associated with, or contributed to, the accident (Stutts 2005).
- **Laboratory based - Simulation studies**, that typically use experimental research designs that allow the investigator to systematically introduce conditions to the driver in a controlled environment and monitor a wide range of performance measures (Strayer 2005).
- **Questionnaires**, that record what people say or think they do.
- **Observational studies**, that focus on what people actually do rather than on what they say they do, examining driving behavior, including driver distraction, using trained observers or in-vehicle electronic recording devices.

2. **Quantifying Driver Distraction**

A special category of observational studies is the investigation of driver distraction using vehicles equipped with various systems recording real driving behavior. They are naturalistic studies in which drivers voluntarily participate in the measurements. The vehicles are equipped with cameras and sensors recording driver distraction at all time. There are two ways to carry out such investigations: in one case, the driver is supplied with an equipment vehicle for a long time and in the second case each participant drives a particular vehicle for a particular route, under the supervision of the researcher. The important advantage of this method is that it is as close as any other method to actual driving. For this reason the results are characterized to be of high validity and reliability. In this study an experimental process in Thessaloniki’s ring road is presented, using an in-vehicle eye tracking system.
2.1 EXPERIMENTAL DESIGN

The equipment used to conduct the experimental part of this study includes:

- The experimental vehicle, model Lancia Ypsilon 1.2 and
- A monitoring system tracking the head and eyes of the driver (Figure 1).

Figure 1: FaceLab monitoring system tracking the head and eyes of the driver

This system consists of:

- Two cameras recording the driver's gaze and head position.
- Infrared lightning to help track the pupil
- Scene camera recording the driver's vision field
- Special chessboard used calibrate the cameras to axes x, y, z
- Data analysis software that automatically correlates the data recorded by the gaze cameras to the data recorded by the cameras that monitor the road in real time.
- Central CPU unit
- Display for adjusting the software and monitoring the driver in real time

The sample for this study was 10 drivers, of the same age group (25-30 years), tested three times on each link of the selected road section. Totally 30 data sets were collected for each direction of the roadway.

All drivers had experience in driving and familiarity with the specific route. They were aware that they were taking part in an experimental testing, but they had no idea what they were being tested about. They just had to drive along the selected roadway just as they would under normal conditions.

The selected area is the eastern part of Thessaloniki’s Ring Road (Figure 2). The Ring Road of Thessaloniki is a six lane (three lanes per direction) urban freeway with a total length of
12.5 km and 12 interchanges. Daily more than 100,000 vehicles, including heavy trucks, travel through the Ring Road and the annual rate of traffic increase is around 5-7% over the last decade.

Speed limit is 90 km/hr but it has been observed that drivers develop speeds greater than permitted. The traffic volume is also high all throughout the day and two VMS signs en route inform drivers about the traffic conditions on the Ring Road. Along the roadside many advertising billboards are setup as well as other elements possibly attracting drivers’ attention.

The experiments were all held during daytime with weather conditions normal to season; not rainy, mostly clear or scarcely cloudy.

2.2 **IMPLEMENTATION**

The procedure in order to conduct each measurement included the following steps:

**Step 1:** The first step was to familiarize the driver with the vehicle and adjust the seat position

**Step 2:** Having settled the driver’s seat, the cameras had to be adjusted to the driver’s height and position in the car, so that the software could identify the gaze and location of the head of the driver.

**Step 3:** Using the chessboard the recording cameras had to be calibrated to the axes of x, y, z.
Step 4: Then, in order to create the head model of each driver images of the face (straight and in angle to the camera) are used on which several characteristic points of the face are set (Figure 4.10) to be used as reference points.

Step 5: Following the steps above, the FaceLab software creates a model of the driver’s head whose position can be identified at any time as well as the direction and movement of the gaze, the direction of the head, the diameter of the pupils and frequency of blinking.

Step 6: The final phase of the experiment was driving the vehicle along the selected area and recording the data.

Figure 3: Recording of driver’s gaze as well as other parameters such as: blinking, fixation, the direction and the duration of the gaze

Figure 4: Final output of FaceLab: a real time representation of the direction of the head and gaze of the driver
2.3 ANALYSIS OF THE RESULTS

For the analysis of results, software Captiv L2100, which is compatible with FaceLab and adequate for video processing, was used. Data was imported in video format and analyzed. The information selected to be analyzed are those related to the direction of gaze in relation to the position of the vehicle characteristics as shown in figure 5 below and in relation to the external environment.

In order to correlate the relation of the recorded gaze direction to the position of the vehicle characteristics a vehicle model was created and imported into the FaceLab (figure 5) model. Moreover a plane in front of the driver, covering the front window, was defined to be the vision field directly related to driving whereas the rest was defined to be not directly related to driving task.

According to the data analysis the driver’s gaze is distributed between the elements of the vehicle and the vehicle exterior environment as shown in the pie below (Figure 6).
The interactions of the gaze and the vehicle characteristics were identified and analyzed towards duration and frequency. During the 9% of the driving time, that the gaze of the driver is directed to the internal part of the vehicle, the distribution among the elements of the vehicle is shown in the diagram below in Figure 7.

It has to be mentioned that not all vehicle characteristics examined are in-vehicle distractions. Some of them such as the rear, left and right mirrors and windows, are made to observe the external environment. However, the analysis provides an indication of how the gaze of the driver is distributed among them.

![Figure 7: Distribution of direction of the gaze on the characteristics of the vehicle](image)

Figure 7: Distribution of direction of the gaze on the characteristics of the vehicle

Regarding the external-to-vehicle distractions, all billboards along the road were identified and mapped for both directions, as well as a special area of the road section on an intersection of the Ring Road, where a great number of illegal posters are set (figure 9). The analysis included a thorough examination of the driver’s behavior, when driving by these potentially distracting elements of the road environment.

![Figure 8: 3D simulation of the head and gaze of the driver](image)
Figure 9: Advertizing billboards recorded along both directions of the specified roadway section

The data analysis pointed out that:

- All drivers are distracted from almost all of the advertizing billboards placed to the roadside of the Ring Road.
- Multiple billboards placed close to one another (figure 10) attract the majority of the drivers.
- Billboards using active elements such as scrolling text or video attracted significantly more and longer glances than conventional static billboards.

Figure 10: Advertizing billboards located by the roadside on a road curve and multiple billboards

- Billboards located in the center or close to the driver’s vision field were more likely to attract the driver’s gaze. For example a billboard placed on the roof of a building attracted 48% of the drivers when it was met along left side of the roadway and 76% of the drivers when it was met at the right side (figure 11).
- Generally billboards placed by the right side of the roadside attract more glances.
Finally driving along the special area of the road section was individually examined making evident that the great number of illegal posters attracts multiple glances from the drivers. Too many small visual distractions at the same level or near intersections and junctions seem to affect the drivers distracting them when intense attention is required.

To sum up, it was made clear that the placement of billboards in the field and the lateral distance from the road is the most important factor for the distraction aroused. Advertizing signs and billboards located in the center of the driver’s vision field, e.g. on intersections tend to get more glances, regardless of the distance from the road.
3. Conclusions

The research conducted so far shows that advertising billboards actually influence driver distraction. However, driver distraction includes lifestyle issues, not just driving issues, including the almost natural tendency of observing objects, events or activities that are new, or attractive. It is therefore of great importance to research driver distraction in order to alarm drivers, improve the law, and even change the design of roads reducing accidents linked to driver distraction.

Even though, their impact on road safety is still not determined, there is overwhelming evidence that, at least in some cases, signs and billboards can be a threat to road safety. Future research should focus on clearly correlating driver distraction to road accidents to adjust appropriately the guidelines of road networks design.

There is apparently a need to correlate driver distraction to specific parameters of advertising billboards and signs (size, brightness, message content, distance from the road) and other characteristics of the driving environment and as well as other in vehicle sources of distraction to eliminate relative crash risks.

Through the experimental procedure carried out, a methodological approach to an investigation of driver distraction was attempted. It emerges from the process that an important and decisive factor for the success of research o experimental design.

It is obvious that this methodology provides a big amount of data that can correlate distraction with various elements of the external environment, of the vehicle but also with driving behavior, all affecting road safety. Much of the analysis of data requires cooperation with experts such as psychologists or doctors in order to provide an integrated approach.

References


Road risk as a presidential priority: a public action for safety and a political instrument

Jacques Chirac was reelected President of the Republic on May 6th 2002. Two months later, on July 14th, during his televised speech, he was questioned by Patrick Poivre d'Arvor, Béatrice Schönberg and Elise Lucet. He answered to the question of main projects of his new mandate: "I would like to mark this five-year period by three big constructions which are not of stone. At first, the fight against the road insecurity. I am absolutely horrified by the fact that the French roads are the most dangerous of Europe". Immediately the media seized this political declaration. During the summer and autumn 2002, decisions rushed and became clearer. September 16th, the President of the Republic convoked the Council of Ministers to the Élysée Palace. He told "to the Government that he expected energetic and obstinate action". September 17th was dedicated to "General States" during which is realized an inventory of fixtures of the road safety in France. The schedule of the political decisions to solve the problem was quickly decided. The deadlines were in short order, what had the effect (wished) to show the strength to this road safety policy. It was announced: "the Government will decide the main measures of his political action during an interministerial committee of road safety [CISR] which will be held within two months. The necessary measures must be taken before the end of the year. A bill on the road safety will be submitted to the Parliament at the beginning of 2003". The interministerial committee of road safety met on December 18th. This day, supervised by the Prime Minister Jean-Pierre Raffarin, were adopted new measures against the road insecurity. The CISR decided to assert the "repressive part" of this fight against the road insecurity by planning to multiply the road controls and to assure of automatic penalties. In a new shape of control of the young drivers, it was also planned to establish a probationary licence valid for three years, according to the idea of Jacques Chirac to cause "a reinforcement of the requirements for the inexperienced drivers". The schedule given by the President has been followed. Parliament’s discussions

---

1 The two other construction were the cancer and the handicap.
2 Press release on September 16th, 2002.
3 In mid-November.
4 Release of the Elysee on September 16th 2002.
took place at the beginning of 2003. It is in the name of Jean-Pierre Raffarin that, on February 26th 2003, Gilles de Robien and Dominique Perben presented the motives for the bill "strengthening the fight against the road violence". That law was adopted then published on June 12th 2003.

Since the presidential declaration of July 14th, 2002 till the decree of application of July 11th 2003, less than one year was needed. One year of political and media excitement concerning road safety.

Various authors wondered about the value of break or continuity of this announcement in the process of structuralization of the devices of fight against the road insecurity. Thus, Samuel Brunet (2004) asked the question: "but is it possible to assert that the road safety went from the shadow into the light?". Rightly, he indicated that "the necessity of punishing the driver is not a recent discovery" and based his words on an analysis of the interministerial committees of road safety realized since 1981. In harmony with this purpose –and by going back in time even farther– we can refer to the history of the traffic control and the penalties decided by the State (Cunegatti, 2009). Because the road safety policy is established for a long time, in particular through a reading of professional bodies asked to assure the existence of the infrastructures and those working to punish the offences to traffic rules, but also by the influence of groups of experts and researchers in road safety who operated " in the end a convergent view of the problem, by condemning the "craft" aspect of control-penalty and the deficit of exhaustiveness of control-penalty existing in France " (Hamelin, 2006).

We add another logic to these points of view. Indeed, this period 2002-2003 saw the strengthening action be centralized by the State. The hypothesis is that particular reasons, strictly political, provoked an acceleration of a firmer road safety policy, with the idea of a strong autonomy of politics. The representatives of the State would have finally taken into account the already existing road security measures but neither really used nor systematized.

This political declaration can be analysed with a triple point of view: the history of the institutionalization of the "road safety" in France, the process of de-individualization of the road risk operated, in particular, by representatives of associations –according to the principle of denunciation developed by Boltanski (1984)– and finally political logics of presidential schedule.

---

5 Parliament’s discussion, n°638, twelfth legislature, registered on February 26th, 2003, concerning "The bill strengthening the fight against the road violence ".

6 Law n°2003-495 of June 12th, 2003, which will be followed by the decree of application n°2003-642 of July 11th, 2003.
A process of progressive transfer on the State of the questions of road safety from the private domain.

Automobile-Club of France – first automobile club born in the world – was created in 1895. At the beginning, driving was an aristocratic activity. At the dawn of the XXth century the road traffic appeared clearly liberal: the drivers behavior is a gentlemen's affair which recovers a “spirit of club” where the art to drive as a lifestyle is a culture, until a decree created in 1921 what will be called the "traffic rules". The control of the road behavior is so officially born through this code which, for the first time, structured the good manners to interact on the road between users, by having for aim to warn the accidents. Two decades later, in 1949, the Road accident prevention (carried by insurance companies) begins to manage the safety on the register of the risk prevention and the education of the drivers. This association became the partner of public authorities by receiving, in 1955, the status of state-approved association. At the same time, and until 1972, the National Union of the Associations of Tourism (UNAT) built the road statistics, the management of the national file of the driving licences and the driving’s exam. Year 1972 indeed sees the State control all these domains connected to the road traffic still detained by the private sector.

A bend\textsuperscript{7} is simultaneously taken when the theme of the road insecurity adopted a national dimension by touching the public opinion thanks to the strong mediatization of two events. The first one came true under the impulse of Christian Gérondeau's proposition, with the decree of July 5th 1972, establishing an Interministerial Committee of the Road safety (CISR) and the creation of a post of interministerial Delegate in the road safety (DISR)\textsuperscript{8}. According to Christian Gérondeau's terms: "I met in the spotlight of media with an intensity which we can’t imagine. When I took up my duties, I had to make ten minutes during the evening information on both television channels! There were only two channels. Thus that means something important. A fame which most of the Ministers never have, that’s evident!" The second event was a very innovative (and very mediatized) action realized by the Road accident prevention. To touch the public opinion, the association, supported by journalists, chose a city of the Tarn where the population was equivalent of the number of victims on roads in 1972 (a little more than 16 500 persons). This called operation "Mazamet, dead city" took place on May 17th 1973. It consisted in showing the city with the bodies of the inhabitants lying on roads and pavements.

Less than ten years later, in 1982, the interministerial permanent team of the road safety was dissolved to put in place a Direction of the Safety and the Road traffic (DSCR) within the Ministry of Transport. This Direction became the objective structure of application of the road safety policy. If the passage by the history shows a frank evolution of the mode of regulation

\textsuperscript{7} Pierre Lannoy (1999), in particular, considers that at the " beginning of the fifties, the question of the traffic became an affair of state ". However it is necessary to distinguish what is linked to the "road traffic" and what must be linked to the different notion of "road safety".

\textsuperscript{8} Christian Gérondeau will take this post and will be the first interministerial delegate.

Hugues Cunegatti - Ifsttar
of the by the State, the next years will again confirm this principle. So the decree of November 24th 1993 organised an Interministerial observatory of the road safety (ONISR), whereas the same year the road education entered the School by establishing a "diploma"called"school certificate of road safety". In 2000, the road safety took the status of "big national cause" and a National Council of road safety is created on August 28th 2001, when Lionel Jospin was Prime Minister.

A process of de-individualization of the road risk, operated by representatives of associations which had the effect of building the problem as a collective problem, until make it perceive as a "national disease".

During two last decades, it is necessary to understand that the role of the State came along with a diversification of the associations of promotion of the road safety appeared in the 1980s and developing during 1990s. The first strong step is, because of an unprecedented mediatization of dramatic accidents, when the militant associations of a new type are born, associations focalised on the defense of the victims and no more on the free use of the road. So the Anne Cellier foundation was established in 1987 by Christiane Cellier, after the death of her daughter, due to the consequences of a car accident. This configuration squares with the model of the denunciation analysed by Luc Boltanski (1984). It is, at first, about a mother revolted by the treatment that she considers inequitable that the justice decided for the driver responsible for the death of his daughter.

We are in the typical scenario of an initiative of denunciation made by a private person rebelling against an injustice imposed concerning one person –a daughter– after an accident with irreparable consequences. As Boltanski finely described it, such action will be perceived as legitimize or, to resume his words, as "normal", under precise conditions which make move the cursor of evaluation of the drama from the individual pole towards the collective pole, the private sphere towards the public domain: "The individuals have [finally] very unevenly, according to the nature of the instruments they used, institutional resources, particularly in a crisis or conflict, to put between them the "distance", in order to manage their relations in an impersonal way, according to an identity legally defined according to rules, by using a general argumentation and by making reference to a collective interest: a relation may be de-individualization, indeed, when each of the individuals in cause can, if necessary, be treated as member of a category for which quite other member of the same category could be substituted without the structure of the relation is modified."

This process of legitimization of the private protest is subordinated to the constitution of a cause abstract and impersonal –thus general and universal– that in this particular case the defended population includes, beyond the effective victims of the road, all the potential
victims which must be protected, or put in situation to never exist. Now in Christiane Cellier's situation, as for the other initiators of movements of the same nature as Geneviève Jungersen (of the League against the road violence, LCVR), we are in the scenario of family dramas where "the instruments of de-individualization" are the rarest, unlike what takes place in the business world for example where we can contact syndicates even to political parties (cf. mobilizations concerning the 24 suicides in Orange).

We understand the premises of the de-individualization in the comments of Christiane Cellier when she remembers itself and tells us, with emotion, the reaction of the driver responsible for the death of her daughter when she went in the workplace of this one (he is insurer) and invites itself in her office to show him Anne's picture, his daughter who’s dead in the accident: " he looked at me with a tired voice and said to me: " Madam Cellier, if you would have been to the same dinner as I, you would have been in the same state. That would have been able as well to arrive at you! " 9 [...] I return to my office. And in my car I ask myself, what I have to do to rock French people?"

In her book Letters to Anne (Cellier, 1995), she writes: "for the first time of my life, I was ashamed to be French. " This urges her to create the Anne Cellier foundation to "make something for you, for all the others, to wake up the country."

The objective constraints which weigh on the process of normalization of these movements of protest act for the defense of a cause of general interest: the road safety.

The road safety which can be defined (Cunegatti, 2009) as an ideal 10 set of manners to conceive the use of the road space, based on the social values that are the respect of "living together" and the life protection of all the users. The road traffic must be perceived as the laboratory of a State which controls an exemplary domain of the life in society. Supposing that, it is advisable to question the relations which these protesters have with public authorities on the one hand, with media on the other hand.

Associations mobilized against the "road violence" maintained an inevitably ambivalent relationship to the State because this one was at the same time the object of their intervention to to act in a way of strengthened repression, and the authority which only was capable of giving them a normality, in other words of bringing an official certification of the transmutation of the initial private drama into a cause of public interest. In other words, these associations asserted themselves simultaneously as pressure group (lobby), by interposed media, to act on the State, and are in the same time in a position of total dependence towards this same State from which they received the justification of their existence.

9 We see how the principle of de-individualization can be also used by the persons responsible for accident, by not explaining a general interest, but a social reality, shared by the largest number of persons.

10 As ideal, it always enters conflict with practical reasons of the users and the logics of the political representatives, who more take into account (and it is the foundation of an internal logic in the politics) the social acceptability of the measures of road safety and the ideal of the notion. This fact explains, as we found it, to the existence of a confrontation between clan of the "realists" and clan of the "idealists" within the National Council of the road safety.

Hugues Cunegatti - Ifsttar
The arrival on presidential schedule of a domain where the mortality is presented as such unacceptable that the mortality is avoidable and the easily assessable results of this policy.

A public policy appears "under the shape of a plan of governmental action in a sector of the society or a geographical space" (Thoenig and Mény, 1989). The public policy of road safety corresponds to this first definition but also to one "built by research" which integrates at the same time an anticipated frame by "the decision-makers" as well as the work of reorganization of the researcher aggregating all which can confine a public policy, at the same time in its contents and its dynamics. The road safety policy (included here from 2002) squares consequently with several domains: the constitution of a set of concrete measures, decisions of police and justice, a general frame of action, aimed publics and definition of results to reach. Considering the institutionalization of objective structures dedicated to the road safety, considering the role of associations and the influence of the media, the integration of the road safety in the governmental work, at instigation of Jacques Chirac, is the outcome of the construction of the road risk as public problem. According to the demonstration made in the "Factory of risks" (Gilbert, 2003), this road risk being now on the political agenda acknowledges the fact that it has been identified as a major public issue. Also, it clearly establishes that road insecurity is a part of the “Society of risk” (Beck, 1991) which "indicates a time in which the negative aspects of the progress determine more and more the nature of the controversies which lead the society".

An analysis of this public policy (renewed or "boosted" by the presidential decision) by sequential analysis (Knoepfel et al, 2001) demonstrates, in our way of thinking, that the solutions to solve the problem of the road insecurity quickly concentrated on a program hardening the penalty to the drivers. The values, as the fight "against the inhumanity" (claimed by Jacques Chirac), the idea "to save lives" and as well " not to accept the unacceptable " (an expression often repeated) had an impact all the more hardly as they had widely been chanted by the associations which had so prepared the legitimacy of this type of speech with a strong power of universalization. The road safety became a cause and, more, it contains itself a moral notion. To use Pierre Bourdieu's idea (1994) when he explains the concept of universalization in *Practical Reasons*, we can consider that, speaking about the idea of "road safety", "no man can deny it openly without denying his own humanity". In this way, road safety is morally indisputable. This brings to the possible implementation of this public policy by the use of the control penalty automated (the "automatic radars") as main application to reduce the number of victims.

A report of the Grand journal (TV news) of September 16th, 2003 reported this idea by broadcasting an interview of Nicolas Sarkozy, then Home Secretary. The presenter of the channel Canal + announced that after "the Council of Ministers, a press conference was
organized in the presence of the Prime Minister, of the Lord Chancellor and of Home Secretary. Is this bill too repressive? Listen to Nicolas Sarkozy's answer: "It is not a question of being repressive to be repressive. It is a question of stopping a scandal specifically French. Why people killed themselves or kill the others on our roads." So, to protect lives, the implemented means are justified if they are effective. In road safety, theoretically, only the end result is important, and, to quote Nicolas Sarkozy's words which are revealing the social stake basing the values of the road safety, to stop that "people killed themselves or kill the others on our roads."

Even if the implementation of the Automated Speed Enforcement (ASE) System was subject to criticisms, the saved lives (anticipation of the result a priori) then (a posteriori) the lives really saved –after the decision of this new policy of control– has transformed this implementation as something untouchable from the point of view of values. Right or left politicians, upholders of the repression or the education, nobody as disapproved this system which contributes to protect the society and lives of its members. 11

Road as laboratory of the citizenship and the State power of control

The phenomenon of legitimization of the accentuation of the control becomes integrated into the "process of civilization" such as Norbert Elias developed this concept. By demonstrating the principle of the process of civilization, Norbert Elias "shows as the slightest gesture has importance and becomes integrated into an evolution which tightens an intimate control of feelings and manners", with result the control of violence and death in society. 12

Dead and injured people on the road shock all the more that these dramas go against this process, in the conception of a calmed society which encourages the euphemisation or the disappearance of the violence. It is not moreover harmless that the use of the expression "road violence" 13 was contributory in this process, by acting as a foil to the acceptance of accidents. Any act defined as violent becomes unbearable, and thus socially unacceptable. Because the political arena has for theoretical principle to embody the most universal morality (righteousness, self-sacrifice, disinterestedness 14, defense of human values), it eventually

11 Idea of the management of the death by the political arena which we find under the feather of Dominique Memmi and Emmanuel Taëb (2009).

12 We find this idea described by Norbert Elias in Society of the individuals. He describes the individual consciousness, such as it is established today, as the result of an historic process. So, against the evidence, there is a cultural relativism of the interiority, the autoconstraint (reserve, emotional control). This evolution is understood as part of a process of civilization, as process of interiorization of the constraints and the expulsion of what the society represses, in the first title of which the homicide.

13 We note the introduction of the expression "road violence" in the journalistic and political usage. The link is evident between this expression built by the League Against the road Violence and the implementation of the law of June 12th, 2003 officially called "Law of fight against the road violence". Chantal Perrichon, president of the League Against the Road Violence indicated me the filiation, and Geneviève Jurgensen, not without pride, explained me that this name of the law of 2003 " was due to the actions of the League against the road violence ."

14 For a precise understanding of the phenomenon, see Frédéric Mollé (2006).

Hugues Cunegatti - Ifsttar
seized the road safety by deciding to take care of it, to control it (on 1972, 1982) and to be made the great defender in 2002. President Jacques Chirac\textsuperscript{15} (as embodiment of the State) took in hand this fight for the road safety and claimed his will to stop the “inhumanity” of being victims of an accident. The bill of "fight against the road violence" had been presented during the Council of Ministers of February 26th 2003. Jacques Chirac had declared that: "this bill has to make understand that there is not impunity anymore concerning road insecurity”\textsuperscript{16}. High values of protection against the insecurity so bring to accept a high level of penalty, with the idea of a "zero tolerance" in response to offenders to the law.

A speech of Gilles de Robien, in a declaration of October 20th 2005 in the Brongniart Palace, during the second general States of the road safety, explains that: many observers doubted about the determination of the Government to change things. The solutions were nevertheless unanimous: "the zero impunity, the end of the privileges, to punish in order to dissuade, the certainty of the penalty", in other words make respect rules! " Nicolas Sarkozy, in 2003 (then Home Secretary) was commenting the governmental decisions concerning the road safety: "nobody can escape the fines. No more privilege\textsuperscript{17} or forgetting. A computer system will make automatic fines and it will be necessary to pay them at once, even in case of contesting ". By the system of automated chain since the report of the infringement of the law up to the establishment of the fine or the offence, the control becomes apparently without defect, and gives the image of an exemplary State, where previously, the phenomenon of the "indulgences" ("polices adopt an autonomous rule which substitutes itself rather often for the legal rule", Pérez-Diaz, 1998) marked a "culture of the arrangement" which had for consequence to discredit the reality of the application of the traffic rules and, in a more general perception, the application of the law. The Automated Speed Enforcement System thus has the effect “to bring under discipline) " polices by giving the image of an exemplary and unrelenting State concerning safety.

It is necessary, from then, to have a reading of the Automated Speed Enforcement System following the point of view of "the instrumentation of the public action" (Lascoumes and Le Galès, 2005) –that is to say "all the problems posed by the choice and the use of tools (techniques, manners to operate, devices) who allow to realize the governmental action. It is a question of understanding the reasons to choose such instrument rather than such the other one, but to envisage also the effects produced by these choices”– to report a road safety policy symbolizing, more widely, a public safety policy.

\textsuperscript{15} He used egularly expressions which entered the idea of "civilization".

\textsuperscript{16} Comment reported in the newspaper Press-ocean of February 27th, 2003. Maybe to go to the symbolism of a hardening of the laws concerning the road safety, it is a magistrate, Rémy Heitz, who became the new interministerial Delegate to the Road safety.

\textsuperscript{17} In 2002, the indulgent practices were still current and were known of all. They were widely described by Claudine Pérez-Diaz (1998).
A pre-organized domain of action allowing to quantify easily the efficiency of a governmental action.

We conceive that Jacques Chirac's choice to make of the fight against the road insecurity a priority was not due to the report of a significant increase of the number of deaths on the road. Between 1972 and 2002 the number of victims had been almost divided by 2.3 in passing from 16,545 to 7,242 persons killed. However, it was possible to realize a significant\(^\text{18}\) improvement and, especially, to know easily the improvements. Indeed, the advantage to take the road safety as presidential project was (besides those presented before) an advantage clearly political liking the existence, for a long time, of structures establishing "the figures of the road mortality ". In particular, for the most recent, an interministerial national Observatory of the road safety (ONISR) was created in 1993.\(^\text{19}\) The strength of the fight against the road insecurity leans on the possibility to estimate precisely this policy. The access to the political schedule\(^\text{20}\) is favored by the fact of having a regular communication of results easily presented by media, thanks to "monthly figures" of the road safety, by comparing them to those of the previous year. These monthly figures working then as systematized appointments with the journalists. This principle, purely political, explains why the Head of State and the government seize the question of the road safety. The road safety policy possesses the advantage to be easily available to the general public and quantifiable. The number of deaths on the road produces a "figure idol", an effect of synecdoche of the part (Volle, 1978): "the synecdoche of the part consists in taking a piece of the whole with the very whole, which strikes the spirit by this part, so that we only see this little piece", having the effect of symbolizing a success of fight against the insecurity in the broad sense. The means of evaluation of this policy by the figures of the road insecurity were consequently easy and complete if we consider that (Leca, 1993) the evaluation is "the activity of accumulating, analysis and interpretation of the information concerning the implementation and the impact of measures preparing to act on a social problem, as well as the preparation of the political schedule that it is about those of the Etats-nations or about those of the local authorities, includes all the problems perceived as calling a public debate, even the intervention of the justifiable political authorities." J.-G. Padoleau (1982).

---

\(^{18}\) See a part of the analysis realized by Cunegatti and Suaud (2008): "the failure of the road safety policy realized between 1997 and 2002 is obvious according to the questioned associative actors. This verdict leads especially since the results fixed by Jean-Claude Gayssot in 1997 was calculated: decrease in five years the number of victims of the road by two. The reality was quite various because the number of deaths decreased only 2.5\% over this period. The report is made by Claude Got: "2.5\% of decline [of the mortality] between May, 1997 and 2002. 2.5\% while Gayssot had announced 50\%! A failure. It is the reality of the failure of the Left on this file. " The subject of road safety can become, by counterpoint, a domain where the Right marks its difference.

\(^{19}\) Even if Jean Chapelon (2007) wrote -to show the improvements- in the "balance sheet 2007 of the road safety in France " (270 pages) that "at the end of 1970s, the balance sheet of the road insecure in France was containing twenty pages with three graphs, twenty three boards and a page of comments ", the analysis of the road risk was already completely quantified and followed, if only by an annual balance sheet, for more than thirty years. Quotation from The road safety in France, Balance sheet of year 2007, The French documentation, 2008, p6. "Word of the editorial staff" by Jean Chapelon, General Secretary of the Interministerial National Observatory of the Road safety (ONISR).

\(^{20}\) "The political schedule that it is about those of the Etats-nations or about those of the local authorities, includes all the problems perceived as calling a public debate, even the intervention of the justifiable political authorities." J.-G. Padoleau (1982).
new measures. " Besides, to quote the comments of Pierre Muller (2008), and with a certainty of efficiency concerning the road safety (contrary to the point carefully presented by the author): " one of the most symbolic stakes in the administrative modernization is doubtless the implementation of procedures of evaluation of public policies even if these policies did not still allow to obtain the anticipated results".

These specificities of the road safety, established during the analysis, were consolidated by the interview realized with Rémy Heitz in 2009\textsuperscript{21}: " it was very important for the Ministers who were associated with, to show that the political voluntarism allowed to refuse this kind of fatalism. The political voluntarism relieved by practical, concrete measures, could give results (...) Very computable, very considerable, very concrete. (...) And I think that played for the success of this project because, and maybe regrettable, we shall remember less the results of the cancer plan and those of the handicap plan because we have this extremely tool [the ONISR] which reacts quickly"

Consequently, the existence of a precise measuring instrument " provided with the prestige of the science " (Desrosières, 2000), "objective" and regular of a risk (because the deaths on the road have the property to be quantifiable) and its capacity to be used by the media thanks to the " strong media interest for this project", (Marchetti, 2008), partially, allowed to make the choice of the fight against the road risk as a political priority.

**Conclusion**

Beyond the indisputable will to go farther in the fight against the road risk (intensification of the influence of the State on the question of the road safety / insecurity), the political choice to choose the road insecurity as " presidential project " also allowed to set up a successful policy – the policy of fight against the road insecurity– this predictable political success acting in a positive way on the image of a government in capacity to act on the security of the persons, this for a period of centring on the problems of public insecurity. It is to the centring of a general policy on the problems of insecurity that we owe the particular treatment –as a priority– concerning the road safety. These multiple games of particular interests become national interest that we can understand the justifiable decision to use the Automated Speed Enforcement System, seen in the public opinion through the image of the "radar". The "success" of this repressive instrument explains, by certain sides, because it "embodies" a nevertheless real and never really clarified idea of the citizen, in other words the subject never totally converted to the general interest and, as such, with difficulty educated to accept the rules of road traffic. The most significant role of associations is perhaps less to convert the French people to the legitimacy of a "pacificated road" that to prepare the opinion for the implementation by the State of a repressive tool by which everybody agrees to

\textsuperscript{21} In July, 2002 Rémy Heitz was councilor for justice for Jean-Pierre Raffarin. He became delegate interministerial in the road safety from 2003 to 2006.

Hugues Cunegatti - Ifsttar
recognize, in an almost magic way, the results in terms of reduction of killed people on the road.

Finally, the analysis of the case of the road safety demonstrates that the knowledge of a risk can be broadcasted and risk be known more easily, even instrumented by politics, when there is a production of an indicator of risk, when the evaluation of the risk is easy and when the profit of this policy is allegedly likely to be a success.


Cunegatti H., C. Suaud (2008), *Associations: what roles and which influences on policies of road safety in France?* Operational group 3 of the national Program of Research and Innovation in the Ground Transport.


Effect of thin water film on tire/road friction

ABSTRACT

Water film on pavement surfaces entails a decrease of friction between the tire and the road. Nevertheless, only effects of water films above 1mm depth were investigated until now, considering hydroplaning risk. From these investigations, formulae were derived to predict the so-called “hydroplaning speed” at which happen hazardous situations for the driver because there is no more contact between the tire and the road. However, a significant number of accidents occurs on very thin water film such as when the road is drying after rainfalls. Indeed, during these periods, drivers have the feeling that the road is dry leading to inadequate behaviors.

This work focuses on the effect of thin water film (< 1 mm) on pavement friction and aims at modeling the friction decrease. Experimental simulations were conducted in controlled conditions in laboratory using the Dynamic Friction Tester machine on various pavement surfaces. Pavement surfaces were wetted by a spray, which was weighted using a precision balance scale in order to estimate precisely the amount of sprayed water. The tire/road lubrication regime can be identified considering the Stribeck curve. A so-called “critical” water thickness, defined as the amount of water from which pavement friction collapses, was determined for each tested pavement surface. Critical water thicknesses are discussed as a function of surface macrotexture, microtexture and test speed.

Keywords: friction, water film, lubrication, pavement surface, texture.
1. **INTRODUCTION**

It is well known that tire/road friction decreases when the road surface is wet. Moore (1975) for example showed that water acts as a lubricant and reduces the fraction of the tire/road contact area where friction forces are generated. Despite this widely accepted explanation, few works have dealt with the relationship between the water depth and the tire/road friction. Based on friction tests using a vehicle equipped with trailer, Veith (1983) showed that the friction coefficient is independent of water depth at low speeds (30 to 50 km/h) but it is strongly influenced by water depth at high speeds (96 km/h or greater). It was found that the friction coefficient varied as the logarithm of water depth (Figure 1). Water film thicknesses greater than 0.12 mm were studied.

![Figure 1: Friction versus water depth for full and half skid depth tires (Veith, 1983)](image)

Models were also published on the relationship between the so-called “hydroplaning speed” and the tire inflation pressure (Moore, 1975; Veith, 1983). The “hydroplaning speed” is defined as the speed limit above which the driver can no more act on his vehicle to control its trajectory (Cerezo et al., 2010). This well-known situation called “hydroplaning” occurs on flooded roads where the water depth is above 1 mm.

Even if the abovementioned works have contributed significantly to the reduction of hydroplaning risk, knowledge is still needed regarding the effect of thin water film on tire/road friction. This situation occurs after rainfalls or during drizzles where the damp aspect of the road surface provides a safety feeling; driving speeds are then as high as those practiced on dry roads. Nevertheless, experimental studies (Delanne et al., 2006) showed that friction at a “damp” state can already be significantly lower than that at a “dry” state (Figure 2). The drastic drop of friction coefficient shown in the figure 1 explains why accident records are generally high after rainfalls. The tire/road contact loss on damp road surfaces is sometimes referred to as “viscoplaning” in order to emphasize the viscous effect of thin water depths.
The single paper dealing with thin water depths is based on works conducted by Kulakowski and Harwood (1990). Using a dedicated laboratory device, they performed friction tests at different water depths and found that the relationship between the friction coefficient and the water depth can be approximated by an exponential function (Figure 3):

\[ \mu = \Delta \mu e^{-\beta h} + \mu_F \]  

(1)

Where
- \( \mu \): friction coefficient;
- \( h \): water depth;
- \( \mu_F \): final friction coefficient;
- \( \Delta \mu \): difference between \( \mu(0) \) (\( \mu \) at \( h = 0 \)) and \( \mu_F \);
- \( b \): parameter of the model.

Actually, Kulakowski and Harwood (1990) supposed that the friction coefficient reaches a level – that is \( \mu_F \) – at which there is no more variation with increasing water depths. Even if experimental conditions are not identical, it can be noticed some similarities between the graphs of figures 2 and 3 respectively.
The same authors defined a critical water depth $h_{crit}$ as the depth at which the dry friction $\mu(0)$ has lost an equivalent of 75% of $\Delta \mu$ (Figure 3). The 75% threshold was chosen arbitrarily. Field tests were conducted at 64 km/h to study the influence of asphalt formulation and tires on induced critical water depth (Kulakowski and Harwood, 1990). The results indicate that critical water depth lies between 0.025 and 0.23 mm for different combinations of pavement surfaces and types of tires. It is important to notice that the tested pavement surfaces are representative of a wide range of roads and results show that, for each test configuration, very thin water film thickness can decrease critically the friction coefficient. The wide range of values of critical water depth points out the significant influence of both surface texture and tire.

2. RESEARCH NEEDS AND METHODOLOGY

From the brief review presented above, it can be said that research on the effect of thin water depths on tire/road friction is still needed to get a more comprehensive understanding of lubricated tire/road contact. In addition to scientific interests, results of this research can be useful for road authorities looking for the way to inform road users about road skid resistance under bad weather conditions. Applications can also be developed by car and tire manufacturers to assist drivers unaware of slippery road. Within the frame of the European project SKIDSAFE (7\textsuperscript{th} Framework Program) dedicated to the modeling of tire/road friction at different scales (materials, tire, vehicle), Ifsttar has initiated a research aiming at developing a model predicting the onset of visco- and hydroplaning from the knowledge of road materials, tire characteristics and tire/road contact conditions (speed, wheel slip, water depth, etc.). The work presented in this paper is part of the Ph.D. carried out by the first author and focused on the viscoplaning aspect. It is composed of two parts:

- in the first part, a dedicated laboratory set up was developed to perform friction tests at different water depths from dry to flooded states;
- in the second part, using experimental results from the first part, definition of a critical water depth was derived in a more physical way than it has been done up to now. Analyses were then done to assess the influence of various test conditions on the critical water depth.

<table>
<thead>
<tr>
<th>Nomenclature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$, $h_0$</td>
<td>parameters of $\mu$ model</td>
</tr>
<tr>
<td>$\eta$</td>
<td>fluid viscosity</td>
</tr>
<tr>
<td>$\mu$</td>
<td>coefficient of friction</td>
</tr>
<tr>
<td>$\mu_F$</td>
<td>coefficient of friction for thick water film</td>
</tr>
<tr>
<td>$h$</td>
<td>water film thickness</td>
</tr>
<tr>
<td>$h_c$</td>
<td>critical water film thickness</td>
</tr>
<tr>
<td>$p$</td>
<td>contact pressure</td>
</tr>
</tbody>
</table>
The most adequate friction-measuring device was first sought for laboratory tests. The device must be compact enough to be easily transportable. In the meantime, it must perform friction tests at different speeds since it is well known that tire/wet road friction is speed dependent. Among the three devices available at Ifsttar (British Pendulum, Wehner/Schulze machine, Dynamic Friction Tester), the last one was preferred. Actually, the British Pendulum is widely used and its results are correlated with braking tests (Giles et al., 1964). However, tests can be performed only at one speed ($\approx 12.5$ km/h). The Wehner/Schulze machine (Do et al., 2007) can provide a friction/speed curve from a full braking test from 100 km/h to complete stop. However, this machine cannot be moved and, due to the set up defined by the machine manufacturer, the water depth cannot be controlled and measured.

The Dynamic Friction Tester (DFT) is widely used in North America, Asia and Oceania. The machine is composed of a measuring unit (Figure 4) and a control unit. The measuring unit consists of a horizontal fly wheel and disc which are driven by a motor. Three rubber sliders are attached to the disc by leaf springs. They are pressed on the test surface by the weight of the device and are loaded to 11.8 N each.

The slider (Figure 5) is a steel backing plate to which is bonded a rubber slider whose size is 6 mm $\times$ 16 mm $\times$ 20 mm. The rubber slider is shaped so as to provide a contact pressure of 0.15 MPa.
A water supply unit is provided to maintain a wet condition of the test surface. Based on the related standard (ASTM, 2009), when the tank is 0.6 m above the measuring unit, a water flow of 3.6 l/min is maintained. If the flow is initiated when the rotation corresponds to a tangential velocity of 85 km/h, the water film thickness will be 1 mm by the time that the velocity reaches 90 km/h and the measurement is initiated.

The test procedure is standardized by ASTM (2009). The test wheel (Figure 4) is accelerated until it reaches a linear speed of 80 km/h. Water is projected on the measured surface by means of two pipes (Figure 4). The motor is then stopped and the test wheel is dropped. When the rubber sliders are in contact with the measured surface, the wheel rotational speed decreases due to the friction generated between the sliders and the surface. Due to the forces on the rubber sliders, displacement occurs in a spring balance. This displacement is then converted to an electrical signal. The speed of rubber sliders is measured from the output of a rotational speed dynamo.

The measurement output is a braking curve from 80 km/h to complete stop; example is shown in the figure 6. Values of friction coefficient, typically at 20, 40 and 60 km/h, are extracted, recorded and displayed the screen of the control unit. Full sliding conditions occur in the contact area between the DFT sliders and the test surface. The slider dimensions as well as the contact condition (full sliding) might then simulate tire tread rubber blocks sliding during a locked-wheel braking. The standard deviation of eight measurements on the same test surface ranged from 0.044 at 30 km/h to 0.038 at 60 km/h (ASTM, 2009).
3.2. Definition and measurement of water depths

There are different ways to define a water depth; two of which are shown in the figure 7 (Veith, 1983).

The water depth above asperity summits can be measured by means of devices equipped with needles (usually two conducting electrodes moving vertically). The water depth corresponds to the height difference between the needles when one touches the measured surface and the other touches the water surface (electrical contact). A major drawback of this method is that the measurement is local and cannot represent the mean water depth on the surface unless a significant number of measurements is done. Non-contact optical water depth sensors can be used too. Unfortunately, water depths less than 1 mm – range of interest for this study – cannot be measured using devices available at Ifsttar. Thereby, a more basic method was chosen to evaluate the wetness of the surface. A spray was used to wet the surface. Before and after each friction measurement, the spray was weighed in order to know the amount of water sprayed on the test surface (Figure 8). Dividing the volume
of water by the wetted area, an average water depth can be calculated. This method takes into account the fact that water can fill voids in case of a rough surface. The derived water depth corresponds then to the second definition given in the figure 7.

The calculated water depth is called the “initial equivalent water depth” as it is the thickness of the water film before the friction test is performed. Actually, when the test wheel is in contact with the surface and spins, the water film thickness becomes non uniform. The determination of actual depths is complex and can be done only by models such as the one developed by Moore (1967). In the present study, the “initial equivalent water depth” is used for further analyses and referred to as, for the sake of simplicity, “water depth”.

![Figure 8: Spray weighing](image)

3.3. Test specimens

Five test surfaces were used. Specimens are 520 mm × 375 mm × 30 mm slabs. Four slabs were produced in laboratory: a very thin asphalt concrete (VTAC) 0/6 (the numbers indicate the size range, in millimeter, of coarse aggregates); a semi-coarse asphalt concrete (SCAC) 0/6; a sand-asphalt and a mosaic composed of river coarse aggregates. Moreover, to study the effect of ageing, another VTAC slab was sawn from an un-trafficked outdoor test track. The VTAC and SCAC are representative of asphalt formulations used for main and secondary roads in France. Asphalt slabs were made using an Ifsttar compactor. The mosaic was made by placing coarse aggregates in a mold (Figure 9) then covering with sand and resin.
3.4. Test procedure

A plastic plate, in which a circle of 345 mm of diameter was cut (Figure 9), is affixed to the slabs to delimit the wetted surface. The edge of the circle was filled with a sealant to avoid water flowing from the measurement area. Finally, the slabs were covered, except on their upper face, by a layer of waterproof material in order to prevent the water from flowing out of the sample. The spraying operation can be considered reasonably accurate with the specimens used for the tests. Actually, when 6 g (sprayed quantity before each friction measurement) is sprayed on 0.1 m² ($\approx \pi \times \left(\frac{0.345}{2}\right)^2$) and, assuming that 10% of this amount of water is sprayed outside the measurement area, it implies an error of only 0.006 mm on the estimated water depth.

For each friction test, new sliders were used to ensure that slider wear does not affect results. The test surface was leveled and free of any contamination. The DFT was placed above the slab using visual markers to ensure that it is always placed at the same location. Compared with the ASTM standard (2009), the test procedure was modified to study the influence of water depth on friction coefficient. Actually, the machine was programmed to run tests with no water from the supply unit; water was then added uniquely by the spray. After a first measurement performed on a perfectly dry surface, the following procedure was repeated 12 times:

1. Slab surface was wetted by nine sprayings ($\approx 6$ g of water);
2. A friction measurement was performed;
3. Spray was weighed.

At the end, a test using the standard procedure for the DFT was performed.
4. RESULTS AND DISCUSSION

4.1. Shape of the friction/water depth curve

Example of friction/water depth curve derived from laboratory test data is shown in the figure 10. The X-axis represents the water depth. The Y-axis represents the DFT friction coefficient ($\mu_{\text{DFT}}$) extracted from the braking curve (Figure 6) at speed $V$ ($V = 20$ km/h in the figure 10).

![Figure 10: Example of the evolution of friction coefficient versus water depth](image)

The observed shape is completely different from that found in previous works (Veith, 1983; Kulakowski and Harwood, 1990). During the first tests (unpublished) aiming at developing an appropriate test procedure, water was sprayed on the primarily dry surface until it appears wet. This procedure induced actually an exponential variation of friction with water depth. The difference between the figure 10 and published results can then be attributed to the water quantity sprayed on the dry test surface to obtain the first wet state. If too much water is sprayed, the transition from “dry” to “wet” can be missed. The shape of the curve derived from our experiments is similar to that of the well-known Stribeck curve (Figure 11) although the X-axis is not the same.
The Stribeck curve relates relative velocity ($V$), fluid viscosity ($\eta$) and pressure ($p$), using the number $\eta V/p$, to friction coefficient. Schipper (1988) proved that there exists a relationship between the lubricant depth ($h$) and $\eta V/p$ in the case of elasto-hydrodynamic lubrication:

$$h \propto \left( \frac{\eta \cdot V}{p} \right)^{0.7}$$

(2)

Based on Schipper results, it can reasonably be said that the obtained friction/water depth curve exhibits the same lubrication regimes as those identified in a Stribeck curve (Figure 11): boundary, mixed and hydrodynamic. It means that the understanding of tire/wet road friction mechanisms, and consequently their modeling, can be enhanced by taking benefit of existing knowledge acquired in tribology. In terms of physical mechanisms for example, if an exponential variation prevails, it means that only mixed and hydrodynamic lubrications occur at the tire/road interface. Based on this observation, the most popular friction/speed model used by road engineers was developed by Leu and Henry (1978):

$$\mu = \mu_0 e^{-bV}$$

(3)

Where $\mu$: friction coefficient;
$V$: relative velocity;
$\mu_0$: friction coefficient at $V = 0$;
$b$: constant.

However, other friction results highlighted the fact that the friction/speed dependency should have rather an inverse-S shape (Mancosu et al., 2000). Unfortunately, the model suggested by the authors of the reference, due to its complexity, is not adopted as it would deserve.

### 4.2. Modeling of the friction/water depth curve
By observing the shape of the curve (Figure 10) and noticing the similarity with respect to the Stribeck curve and works on water depth, a model was developed to fit the shape of the friction/water depth curve derived from our experiments:

\[
\mu = \Delta \mu \cdot e^{-\left(\frac{h}{h_0}\right)^\alpha} + \mu_F
\]

Where
- \(\mu\): friction coefficient;
- \(h\): water depth;
- \(\mu_F\): final friction coefficient;
- \(\Delta \mu\): difference between \(\mu(0)\) (\(\mu\) at \(h = 0\)) and \(\mu_F\);
- \(h_0, \alpha\): constants.

The new model is similar to that proposed by Kulakowski (1990) but can simulate other shapes than the exponential one. Actually, if \(\alpha = 1\), an exponential variation can be found. For \(\alpha \neq 1\), other shapes can be found. Figure 12 shows how the model (4) fits experimental data.

The word “critical” is used by road authorities to decide whether warnings must be sent to road users or not. The same word can be used by car manufacturers to activate driver assistance systems. In the context of warning or driver assistance on wet roads, people look for a critical water depth, which can be measured or estimated in real time, above which something must be done. Based on the shape of the friction/water depth curve, it appears that the most critical phase is the transition from boundary to mixed lubrication regimes where friction can drop drastically even if the road surface is still apparently “safe”. Dividing the friction/water depth curve into three parts: boundary, mixed and hydrodynamic lubrication, each part of the curve is linearized as shown in the figure 13: a horizontal to represent the stable friction level during the boundary lubrication phase, a sloped line to represent the mixed lubrication phase, and again a horizontal to represent the hydrodynamic lubrication phase. Actually, friction should increase slightly during the hydrodynamic phase due to drag.

**Figure 12: Fit of experimental friction results by the new model**

### 4.3. Physical definition of a critical water depth

The word “critical” is used by road authorities to decide whether warnings must be sent to road users or not. The same word can be used by car manufacturers to activate driver assistance systems. In the context of warning or driver assistance on wet roads, people look for a critical water depth, which can be measured or estimated in real time, above which something must be done. Based on the shape of the friction/water depth curve, it appears that the most critical phase is the transition from boundary to mixed lubrication regimes where friction can drop drastically even if the road surface is still apparently “safe”. Dividing the friction/water depth curve into three parts: boundary, mixed and hydrodynamic lubrication, each part of the curve is linearized as shown in the figure 13: a horizontal to represent the stable friction level during the boundary lubrication phase, a sloped line to represent the mixed lubrication phase, and again a horizontal to represent the hydrodynamic lubrication phase. Actually, friction should increase slightly during the hydrodynamic phase due to drag.
forces. However, within the experimental set up, no variation curve experienced such a tendency; a horizontal constitutes then a good approximation.

The critical water depth is then defined as the amount of water obtained at the intersection between the boundary lubrication line and the mixed lubrication line (arrow in the figure 13). From this point, a small additional amount of water is sufficient to degrade significantly the friction coefficient. During the experiments, critical water depths lesser than 1 mm were found. Such critical water depths can also determine the onset of viscoplaning as discussed in the introduction.

The remaining of the paper focuses on the critical water depth and highlights how it is affected by factors related to the tire/road contact such as speed or road surface roughness.

4.4. Influence of the test speed

Figure 14 shows the variation of the friction/water depth curve at 4 different test speeds. The very thin asphalt concrete specimen (VTAC) is considered. The critical water depth was determined for each graph.
The critical water depth tends to decrease with test speed (0.34 for 20 km/h, 0.31 for 40 km/h and 0.15 for 60 km/h). This tendency is logical since increasing speed leaves less time to evacuate water from the tire/road contact area and, consequently, induces a rapid transition to the mixed lubrication regime. Everything happens as if a more important water quantity was sprayed on the dry road due to increasing speed. This explanation is confirmed by the fact that at 70 km/h, the first stage of the curve is not visible and the friction/water depth curve becomes exponential.

Figure 15 summarizes the evolution of critical water depths versus speed for all the specimens. For each specimen, two series of tests were conducted. It can be seen that each specimen exhibits the same tendency: critical water depth values obtained at 20 and 40 km/h are the same and higher than that at 60 km/h. It suggests that the effect of speed on the critical water depth is only significant above a certain speed. This confirms observations made by Veith (1983).
4.5. Influence of the road surface macrotexture

Road surface texture can be divided into different scales defined by the wavelength and the peak-to-peak amplitude of its components. The scales influencing pavement friction are macrotexture and microtexture. They are conventionally defined as follows (PIARC, 1987):

- **Macrotexture**, for wavelengths between 0.5 and 50 mm and peak-to-peak amplitudes from 0.1 to 20 mm, is the surface roughness quality defined by the mixture properties of an asphalt material. It plays an important part in evacuating the bulk water.

- **Microtexture**, for wavelengths under 0.5 mm and peak-to-peak amplitudes from 1 to 500 µm, is the surface roughness quality at microscopic level. It is a function of the surface properties of the aggregate particles contained in the asphalt. It assists in squeezing thin water films in order to provide effective contact between road and tire.

In order to study the influence of macrotexture, the root-mean-square deviation of a surface profile $R_q$ was calculated. It is given by the following formula (Stout, 1993):

$$
R_q = \sqrt{\frac{1}{N} \sum_{i=1}^{N} z^2(x_i)}
$$

(5)

Where $z(x_i)$: surface profile height at abscissa $x_i$; $N$: total number of surface profile points.

The profile $z(x)$ is obtained by subtracting the least squares datum plane from the original surface. $R_q$ is a parameter which gives a good evaluation of the macrotexture of the surface.

Table 1 lists $R_q$ values obtained for all the specimens. The measurement of the surface profiles was done by means of a non-contact laser profile meter (15 profiles by sample; profile length: 76 mm; x-resolution: 10 µm).
Table 1. Values of $R_q$ for the test specimens

<table>
<thead>
<tr>
<th>Specimen</th>
<th>$R_q$ (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VTAC 0/6</td>
<td>0.84</td>
</tr>
<tr>
<td>Aged VTAC 0/6</td>
<td>0.72</td>
</tr>
<tr>
<td>SCAC 0/6</td>
<td>0.55</td>
</tr>
<tr>
<td>Sand-asphalt</td>
<td>0.72</td>
</tr>
<tr>
<td>Mosaic</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Figure 16 shows the evolution of critical water depth versus $R_q$ for three different test speeds. A linear relationship between these two parameters can be observed. The observed tendency is logical: surfaces with a “good” macrotexture (high $R_q$) evacuates water from the tire/road contact area better and postpone the transition to the mixed lubrication regime.

![Graph showing evolution of critical water depth versus $R_q$ for three speeds](image)

Figure 16: Evolution of critical water depth versus root-mean-square deviation of the surface for three speeds

4.6. Influence of the surface microtexture

Since it is difficult to make samples, using natural materials, with the same macrotexture and different microtextures, the mosaic and the aged VTAC, both having similar $R_q$ values, were compared as an attempt to assess the influence of the microtexture. In fact, unlike the aged VTAC, the river coarse aggregates from the mosaic have no microtexture.
Figure 17 shows that, while the friction coefficients during the boundary lubrication regime are similar, the mosaic experienced a more drastic friction drop as soon as the lubrication regime becomes mixed. It can also be noticed that the critical water depth does not seem to depend on the microtexture. It is almost surprising since, looking at the order of magnitude of microtexture roughness, one can expect that thin water depths are related to the surface microtexture.

Nevertheless, the figure 17 confirms the role of microtexture on tire/road friction. When there is no water between the specimen surface and the slider, the contact area – therefore the coefficient of friction – is indeed greater on a smooth surface than on a rough one. It can be clearly seen in the figure 17 that the first red point (mosaic, without microtexture) is higher ($\mu_{DFT} \approx 1.25$) than the first blue point (aged VTAC, with microtexture) ($\mu_{DFT} \approx 1$). However, when the surface is wetted, asperities, mainly sharp ones, are needed to squeeze out the water film. Otherwise, as it happened with the mosaic specimen, water, even at very thin depths, can penetrate very quickly and causes contact loss between the tire and the road.

The above analysis is a first attempt to assess the influence of microtexture on the onset of viscoplaning. A more rigorous analysis will be done using more quantitative microtexture descriptors derived from surface profiles.

![Figure 17: Influence of microtexture surface on coefficient of friction](image)

5. **CONCLUSIONS**

In this paper, study of tire/road lubricated friction was presented. Focus was made on the effect of thin water films (< 1 mm) which occur after rainfalls or during drizzles. Despite the apparent safety feeling provided by a damp aspect of the road surface, thin water films can alter already significantly the available road friction and reduces contact between the tire and the road.

In the first part, the development of an experimental set up was presented. Using a rather basic method to wet the surface and estimate the average water depth, friction/water depth curves were obtained. Different in shape from other previously published curves, the new
curves are unexpectedly similar to the well-known Stribeck curve. It was then possible to identify different lubrication regimes occurring at the tire/road interface. Using a simple mathematical function to represent the observed results, a so-called “critical water depth” was derived. It represents the water depth at which transition between the boundary and the mixed lubrication regimes occurs. This new definition, compared with the few ones found in existing literature, is more physical and can be used to determine the onset of viscoplanning.

Further analyses were made to assess the influence of some factors on the critical water depth. It was found that critical water depth decreases with speed and increases with the road surface macrotexture. Both observations are logical and can be explained by considering the way by which water penetrates the tire/road contact area. The influence of road surface microtexture is more unexpected and needs further investigations.

Acknowledgements

The author gratefully thanks the support of the SKIDSAFE project (Enhanced Driver Safety due to Improved Skid Resistance) of the FP7 Surface Transport Programme. Mr. Patrick Maisonneuve is acknowledged for his technical support on using the Dynamic Friction Tester.

References

Mancosu, F., A. Parry and F., La Torre (2000). Friction Variation Due to Speed and Water Depth, 4th International Symposium on Surface Characteristics SURF, Nantes, France.
THE DEVELOPMENT OF ROADSIDE SAFETY CRITERIA FOR PORTUGUESE ROADS

1. INTRODUCTION

Each year, almost 40,000 persons die and about 1.6 millions are injured in the European Union, as a result of road accidents. According to estimates, the damages so incurred are almost 2% of the EU’s NGP. The latest monetary estimates for road accident losses in Portugal date back to 1995, when they amounted to over 3100 million euros (3% of the country’s NGP).

Accidents involving roadside dangerous objects make an important contribution to the total number of road injuries. From 2003 to 2007, ran-off-the-road accidents (RORA) accounted for 29% of the total number of injury accidents in Portugal (37% of the total number of fatalities). In addition, Portugal has, perhaps a paradoxical evolution in this matter. The fact is that the number of RORA accidents has consistently increased, although the total number of accidents (which includes RORA) has been declining continuously (Figure 1).

![Figure 1 - Development in the numbers of accidents and RORA in Portugal (1991-2006).](image)

Crash data indicate that roadside geometry, including slopes, embankments, and ditches, contributes to more than half of all run-off-road accidents involving serious injury or death.
These roadside features are believed to be the leading cause of rollover in single-vehicle RORA.

Results of recent studies made at Laboratório Nacional de Engenharia Civil – LNEC (National Laboratory for Civil Engineering) have shown that important differences may exist between RORA in some Portuguese roads, when compared with RORA in roads from other countries. Indeed, a higher percentage of collisions with obstacles in the inside of horizontal curves were registered in Portugal, than in other published studies of safety in road curves (Cardoso, 2007).

According to the European project RISER (RISER, 2006), improved data collection (based on real world crash information) combined with the establishment of guidelines for design (founded on European best practices) can reduce the total number of collisions as well as mitigate the severity of those collisions that occur.

Several studies made all over the world point out to the convenience for the analysis of roadside safety features according to the following perspectives (Corben B. et al., 1997; Highways Agency, 2002; Mak K.K., Sicking D.L., 2003; SETRA, 1999; SWOV, 2007):

- Characteristics of the most common accidents (speed, encroachment angles, vehicle types, encroachment rates, vehicle swath, etc.).
- Characteristics of roadside hazards at accident sites (length, width, location, severity, crashworthiness, etc.).
- Crash costs.
- Characteristics of roadside features at accident sites (road restraint systems, shoulders, slopes and medians), and the resulting benefit/cost ratios.

The analysis of roadside safety features heavily depends on the specific characteristics of each country’s road network, car park and drivers’ behaviour. Furthermore, in-service evaluation of the performance of restraint systems should be performed. This paper provides an outline of the procedure being developed to support decisions concerning roadside safety design in Portugal and selection and installation of restraint systems according to the European Committee for Standardization standards.

2. CURRENT PRACTICE

Roadside design consists in the definition of characteristics of the area between the carriageway edges and roadway right-of-way limits and is an important component of total road design. Concern with roadside characteristics and their influence in road safety is not a new issue. Several studies were made in the USA, since the 1960’s. An evidence of the
Concern in USA regarding this problem is the existence of the “Committee on Roadside Safety Features” in Transportation Research Board, which belongs “Group 2 – Design and Construction of Transportation Facilities”. It has more than 25 members working exclusively on the roadside safety issues.

In Europe, studies on roadside safety were published in The Netherlands and in Sweden in the 1970’s. In the last decade European safety research has paid increasing attention to roadside issues, namely through national and pan European efforts (the SAFESTAR Project, under EU’s 4th FR&DP; RISER, under the 5th; RANKERS, under the 6th; and EuroRap). Also, increasingly the concepts of self-explaining and forgiving roads (RIPCORD–ISEREST project) have been adopted in road design standards, as a means to reduce the risk of driver error and to mitigate the consequences of errors that may still occur, including errant vehicles leaving the roadway and entering the roadside area.

The influence of roadside characteristics on safety was recently analysed in Portugal, at the LNEC, resulting in the publication of a report titled “Rural Roadside Characteristics and Road accidents” (Cardoso, J.L. and Roque, C.A., 2001). Roadside safety features were analysed and a first approach to the economic safety assessment of alternative roadside conditions, for given highway and traffic conditions attempted.

Few European countries have adopted a well-structured and systematic approach towards collecting required information on collisions with roadside obstacles. Yet, without quantitative data and research based operational guidelines, attempts to significantly improve roadside safety will meet limited success.

2.1. Forgiving roadside and clear zone concepts

The general nature of a RORA is that a vehicle will run off the road into the roadside and either overturns or collides with a non-traversable obstacle or with the terrain. Therefore, one of the main factors which determine the severity of these types of accidents is the layout of roadside and the type of objects present which potentially could be hit by errant vehicles.

There are numerous reasons why a vehicle may leave the carriageway. Regardless of the reason, a forgiving roadside can reduce the consequences of leaving the roadway. The ideal road has roadsides and a median area that are flat and unobstructed by hazards. Roadside environment should not contain dangerous elements that will seriously injure or kill vehicle occupants that have unplanned trajectories off the carriageway. A fundamental component of the “forgiving roadside” concept is the definition of an obstacle-free safety zone besides the carriageway. Since this is economically and functionally not always achievable, it is sometimes necessary the installation of passive safety equipment like road restraint and...
energy absorbing systems (or break-away posts) to protect the occupants of vehicles and minimize the consequences from dangerous impact hazards.

It is not easy to find in Portugal examples of roadways which possess forgiving roadsides. As an illustrative example, Figure 2 shows a road section with wide hard shoulders, with gentle side slopes and a metal safety barrier protecting from the sign supports (Roque and Cardoso, 2010a).

![Figure 2- Portuguese road with forgiving roadside.](image)

It is important to recognize that road restraint systems placed near a travel lane are, in fact, themselves potential impact hazards. The proper engineering design of passive safety infrastructure should ensure that any subsequent impact with a safety device is much less severe than the resulting impact if the safety device had been absent.

Thus, clear zones are one of the most important contributors to roadside safety design. A clear zone is “an area free of fixed objects or dangerous slopes, adjacent to the roadway, providing a recovery zone for vehicles that have left the carriageway” (AASHTO, 2002). The principle of the clear zone is based on US research from the 1960s which indicates that drivers of over 85% of vehicles that ran-off-the-road on highways can recover control of their vehicle within nine meters of the edge of the roadway. Guidelines therefore where introduced requiring all roadside hazards to be at a minimum distance, ranging from three up to nine meters from the edge of the road. The required distance is primarily a function of the daily volume of traffic and the 85th percentile speed of vehicles on that section of road.

There is a significant agreement between European countries in terms of general practices. For example, many countries promote the safety zone and refer to European Standards (CEN-EN) to describe the performance of safety equipment. The general approach to roadside safety improvement is also largely the same among countries. However the technical specification of each roadside safety element varies among countries (national dimensions for the clear zone
are presented in Table 1. This variation is the result of differences in the benefit-cost ratios of each solution, further augmented by differences in prevailing national crash characteristics.

Table 1. National dimensions for the clear zone (Highways Agency, 2002; RISER, 2003).

<table>
<thead>
<tr>
<th>Country</th>
<th>Width of clear zone (m)</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>6</td>
<td>For Motorways (10 m if a “dangerous zone”)</td>
</tr>
<tr>
<td></td>
<td>4.5</td>
<td>For Express Roads (7.5 m if a “dangerous zone”)</td>
</tr>
<tr>
<td>Belgium</td>
<td>4.5</td>
<td>For Motorways</td>
</tr>
<tr>
<td></td>
<td>3.75</td>
<td>For Express Roads</td>
</tr>
<tr>
<td>Denmark</td>
<td>9</td>
<td>For Motorways and Express Roads with operating speed ≥ 90km/h</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>For Express Roads with operating speed &lt;90 km/h</td>
</tr>
<tr>
<td>Spain</td>
<td>2.5</td>
<td>For speed limit ≥ 60 km/h</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>For single carriageways with speed limit &lt; 60 km/h</td>
</tr>
<tr>
<td>USA</td>
<td>9</td>
<td>For speed limit of 120 km/h</td>
</tr>
<tr>
<td>Finland</td>
<td>7</td>
<td>For Motorways</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>For Express Roads</td>
</tr>
<tr>
<td>France</td>
<td>10</td>
<td>For Motorways</td>
</tr>
<tr>
<td></td>
<td>8.5</td>
<td>For Express Roads</td>
</tr>
<tr>
<td>Greece</td>
<td>9</td>
<td>For Motorways and Express Roads (19 m near railroads)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>13</td>
<td>If operating speed = 120 km/h</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>For Motorways Roads with operating speed &lt; 120 km/h</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>For Motorways</td>
</tr>
<tr>
<td>Hungary</td>
<td>2.5</td>
<td>-</td>
</tr>
<tr>
<td>Ireland</td>
<td>10</td>
<td>For roads with design speed of 120km/h</td>
</tr>
<tr>
<td>Norway</td>
<td>6</td>
<td>If ADT ≥ 15 000</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>For Express Roads if ADT is “high”</td>
</tr>
<tr>
<td>Poland</td>
<td>3.5</td>
<td>-</td>
</tr>
<tr>
<td>Portugal</td>
<td>3.5</td>
<td>For Motorways and Express Roads</td>
</tr>
<tr>
<td>UK</td>
<td>4.5</td>
<td>For dual or single carriageways with speed limit &gt; 80 km/h</td>
</tr>
<tr>
<td></td>
<td>3.5</td>
<td>For single carriageways with speed limit ≤ 80 km/h</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>4.5</td>
<td>For Motorways and Express Roads</td>
</tr>
<tr>
<td>Sweden</td>
<td>10</td>
<td>If operating speed = 110 km/h</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>If operating speed = 90 km/h</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>If operating speed = 70 km/h</td>
</tr>
<tr>
<td>Switzerland</td>
<td>12.5</td>
<td>For Motorways</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>For Express Roads</td>
</tr>
</tbody>
</table>

2.2. Road restraint systems

Slopes, ditches and trees are common roadside hazards. Yet there is no clear consensus within the international community and amongst the Portuguese engineers on how traffic should be protected against these hazards to guarantee sustainable road user safety levels.
Road restraint systems are roadside structures that include safety barriers, crash cushions, terminal of barriers, the transitions among different road restraint systems, motorcyclist protection devices, etc. Road restraint systems are used to protect vehicle occupants from the roadside environment and are one a major issue in roadside safety. The underlying requirement is that the restraint system will result in a collision severity that is lower than a collision with the roadside obstacle being shielded.

European functional requirements for road restraint systems are covered in the European Standard EN 1317 (European Committee for Standardization, 2007a), using criteria such as containment level, working width and several bio-dynamic indicators. Each European country adopted a full range of national recommendations for the installation of restraint systems based on several criteria (see Table 2).

Furthermore, it should be noted that in Portugal the parameters listed only relate to the installation criteria. In current guidelines there is no reference to the containment level and Acceleration Severity Index (ASI) to consider.

Table 2. National criteria for installation and selection of restraint systems (Roque and Cardoso, 2010b)

<table>
<thead>
<tr>
<th>Country</th>
<th>Speed</th>
<th>Economic analysis</th>
<th>Traffic characteristics</th>
<th>Surrounding characteristics</th>
<th>Side Slope</th>
<th>Roadside obstacles</th>
<th>Weather</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Spain</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>France</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Netherlands</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Ireland</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Italy</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Norway</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Portugal</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>UK</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Sweden</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>South Africa</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Australia</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Canada</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>USA</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>New Zealand</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
</tbody>
</table>

(a) Metallic safety barriers only. (b) According to JAE (JAE, 1994)

It is also noteworthy that differences between countries are small as relates to the minimum and maximum containment levels considered, when compared at an international level. The containment level is an indication of the magnitude of vehicle impact that a safety restraint system can withstand without failure. EN1317 provides a variety of vehicle crash tests with several vehicle types, masses, speeds and impact angles. The higher the containment level the
greater the barrier’s ability to restrain and redirect an errant vehicle. EN1317 defines four main levels of containment (T – temporary, N – normal, H – high, H4 – very high).

Table 3 presents these values for all countries surveyed in the present study, according to European Standard EN 1317 and American Standards NCHRP 350 (National Cooperative Highway Research Program, 1993). In addition the same information is presented converting the values from countries that adopt the NCHRP 350 standard to the classification of European Standard EN 1317.

Table 3. Containment level selection (Roque and Cardoso, 2010b)

<table>
<thead>
<tr>
<th>Country</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>N2</td>
<td>H4b</td>
</tr>
<tr>
<td>Spain</td>
<td>N2</td>
<td>H4</td>
</tr>
<tr>
<td>France</td>
<td>N1</td>
<td>H4</td>
</tr>
<tr>
<td>Netherlands</td>
<td>T1</td>
<td>H2</td>
</tr>
<tr>
<td>Ireland</td>
<td>N2</td>
<td>H2</td>
</tr>
<tr>
<td>Italy</td>
<td>N1</td>
<td>H4</td>
</tr>
<tr>
<td>Norway</td>
<td>N1</td>
<td>H4</td>
</tr>
<tr>
<td>UK</td>
<td>N2</td>
<td>H4</td>
</tr>
<tr>
<td>Sweden</td>
<td>N1</td>
<td>H4</td>
</tr>
<tr>
<td>Australia</td>
<td>TL2 (&lt; N1)</td>
<td>TL6 (H4b)</td>
</tr>
<tr>
<td>USA</td>
<td>TL2 (&lt; N1)</td>
<td>TL5 (H4b)</td>
</tr>
<tr>
<td>New Zealand</td>
<td>TL3 (&lt; N2)</td>
<td>TL6 (H4b)</td>
</tr>
</tbody>
</table>

(a) Metallic safety barriers only.
(b) Could be higher in exceptional condition.
(c) Could be lower in temporary situations (T1 to T3)
(d) Corrected value.

Note: In some countries H4a and H4b levels appear only as H4 because they have no hierarchical relationship.

Finally, it is important to highlight that, in almost all countries analyzed, the installation of very high containment levels (H4a or H4b) is required only in very special situations.

3. DEVELOPMENT OF THE NEW PROCEDURE

3.1. Introduction

In a scenario of growing financial restrictions, it is recommendable that the Portuguese road system should be analysed in order to develop a cost-effective framework for roadside safety decisions, based on the observation of data registered in Portugal and on the in-service performance of installed equipments.

That is the context in which LNEC started the SAFESIDE - Roadside safety research project. Its main objective is to develop a method for assessing the influence of roadside characteristics in Portuguese road safety, which may be used to support roadside safety
decisions concerning the design of new roads and the redesign and operation of existing roads.

The main focus of research activity is concerned with the development of statistical models for predicting roadside accident frequency and severity, fitted to general accident data from the Traffic Authority, and the identification of prevailing characteristics of RORA through analysis of in-depth data collected in pilot road sections.

The roadside safety evaluation model, which consists of a model for the simulation and evaluation of the safety effects of alternative roadside scenarios, will be based on the mentioned accident model (that results in accident costs and number and type of target accidents expected to occur per year) and on a list of roadside safety measures (including their safety effects and the present value of their implementation costs). In order to estimate the number of accidents that can be expected to be prevented per year, two components are expected to be calculated: the number of target accidents expected to occur per year, and the safety effect of the safety measure on target accidents (Figure 3).

The proposed procedure to achieve this goal uses three different complementary national data sources, accident data, roadside and obstacles characteristics and road traffic counts, which are managed by different organizations.

Therefore four steps are involved in determining the recommendation for an efficient roadside safety measure:

1. For a given roadside safety measure, calculate the average number of victims for each injury level per accidents prevented;
2. Multiply the average number of victims by the average costs for casualties avoided for each injury severity level to determine the present value of the benefits;
3. Determine the present value of implementation costs for that measure;
4. Calculate the benefit-cost ratio of the roadside safety measure and compare it with alternative roadside scenarios.

The procedure is intended to support decisions concerning both the design of roadside (remove obstacles, reduce the probability of obstacles being hit, reduce the potential danger of an obstacle, or protect traffic by means of a restraint system) and the selection of the restraint system.
3.2. Accident model

In accident analysis there is a large consensus that poisson and negative binomial regression count models are an appropriate methodological technique for modelling accident frequency (Washington et al., 2003). To illustrate the family of count-model alternatives as applied to accident frequency (number of accidents on a roadway element in some time period) it is used the more popular method of the two: the Poisson regression model (Washington et al., 2003). In applying Poisson regression in accident frequency analysis, let \( n_{ij} \) be the number of RORA on roadway section \( i \) during period \( j \). The Poisson model is,

\[
P(y_{ij}) = \frac{\exp(-\lambda_{ij}) \lambda_{ij}^{y_{ij}}}{y_{ij}!}, \tag{eq.1}
\]

where \( P(y_{ij}) \) is the probability of \( y \) accidents occurring on a highway element \( i \) in time period \( j \) and \( \lambda_{ij} \) is the expected value of \( y_{ij} \),

\[
E(y_{ij}) = \lambda_{ij} = \exp(\beta X_{ij}), \tag{eq.2}
\]

for a roadway section \( i \) in time period \( j \), \( \beta \) is the vector of parameters to be estimated and \( X_{ij} \) is a vector of explanatory variables describing roadway section geometric characteristics,
environmental characteristics and other relevant roadside features that may affect accident frequency.

These models do not explain the average number of participants in each accident nor crash severity - fatal, severe, or only slight injury. However, it is possible to obtain reasonable approximations disaggregating accidents by analogy with historical data by type of road.

### 3.2.1. Accident costs

According to Elvik *et al* (2009) the accident costs are the sum of five main items: Medical costs; loss of production capacity; costs of property damage; administrative costs and economic valuation of lost quality of life.

There are no official values for accident costs in Portugal. The reference values currently used by way of Portuguese economic impact assessments are based on HEATCO (2006) (which are presented as values for casualties avoided). Given in 2010 prices, the Portuguese unit costs per case of injury are stated in Table 4. These values were updated in accordance with official inflation data (Source: Instituto Nacional de Estatística).

<table>
<thead>
<tr>
<th></th>
<th>Fatality</th>
<th>Severe injury</th>
<th>Slight injury</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>860 598</td>
<td>111 996</td>
<td>8 606</td>
</tr>
</tbody>
</table>

### 3.2.2. In-service observation in pilot roads

The accident analysis is being done using the road accidents database from LNEC’s Planning Traffic and Road Safety Division. This database contains information collected by the Police (*Guarda Nacional Republicana* and *Polícia de Segurança Pública*) and treated by the Traffic Authority (*Autoridade Nacional de Segurança Rodoviária* - ANSR).

Information provided to LNEC database does not describe all accident details relevant for roadside safety analysis. In most cases existing information has to be complemented with data obtained through the computational reconstruction of crash events. In this way, several parameters of interest may be calculated, such as impact speed, departure angle and crash pulse.

To collect this information, an in-service observation campaign in selected pilot roads was defined. The identification of the most frequent types of roadside accidents per road category considered in the pilot study and a collection of the characteristics of vehicles involved in those accidents, their manoeuvres, roadside characteristics and type of obstacle hit are the
main objectives of this study. Measurements of traffic characteristics in pilot roads are being made and accidents involving roadside will be reconstructed.

Contacts have been made with private motorway concessionaires, to ensure their support in data collection on pilot motorways stretches. Active cooperation between the research institute (LNEC) and major public stakeholders was also secured, namely: the National Road Administration (Estradas de Portugal - EP), the Traffic Authority and the Police. General accident data is available from the Traffic Authority; additional accident data being provided by the Police; urban and road network data being provided by National Road Administration or other road administrations (municipalities and motorway concessionaires).

As statistical models for predicting accident frequency and severity have not yet been defined, preliminary results based on cross section analysis of Portuguese accident data were calculated for all sections of the National Road Network, where data on traffic volumes is available (in 2004-2007 period). Three rates were defined, as related to the severity of accidents:

- Mortality rate (MR), concerning the number of deaths per $10^6$ vehicles×km;
- Serious injury rate (SEIR), concerning the number of persons seriously injured per $10^6$ vehicles×km;
- Slight injury rate (SLIR), concerning the number of persons slightly injured per $10^6$ vehicles×km;

The average values of those accident rates in 1142 single and 753 dual carriageway road sections are presented in Table 5.

Table 5. Average number of deaths, seriously injured and slightly injured injured per $10^6$ vehicles×km

<table>
<thead>
<tr>
<th></th>
<th>MR</th>
<th>SEIR</th>
<th>SLIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single carriageway</td>
<td>0.007</td>
<td>0.021</td>
<td>0.204</td>
</tr>
<tr>
<td>Dual carriageway</td>
<td>0.003</td>
<td>0.011</td>
<td>0.136</td>
</tr>
</tbody>
</table>

3.2.3. Accident reconstruction results

The reconstruction tool used at LNEC is the Human-Vehicle-Environment (HVE) software, developed by Engineering Dynamics Corporation (EDC). This is a powerful program for the simulation of motor vehicle dynamics in many different accident situations. In this study it is being used to estimate ran-off-the-road vehicle trajectories and impact speeds, based on evidence collected by police officers at the accident scenes.
The HVE analysis tool is being used for the reconstruction of accidents, allowing for a description of the impact conditions and also of the vehicle kinematics during the impact event itself. It takes into account vehicle manoeuvres prior to the crash, and it is convenient for accidents which involved elements of the road environment with complex geometry such as embankments. Accident reconstructions will be used to collect important parameters that are not included in the original accident database and that may be used as transformation factors in the process of estimating the safety effects of roadside intervention scenarios.

3.3. Roadside safety measures

3.3.1. Safety effects

The evaluation of the effects of measures intends to estimate, quantitatively, the effect of performing a corrective intervention on the expected frequency of accidents or casualties. The quantification of the effects of measures aimed at reducing crashes represents a critical point for the application of the CBA techniques to road safety.

The effects of a measure are understood as any change in social welfare (positive or negative) that is the result of that measure (intended or not) (ETSC, 2003). Road safety measures may have impact in three aspects of the traffic system: safety, mobility, and environmental. The aim of the measure is to decrease the damage caused by road accidents. Therefore the main effects to take into account first are the safety effects: the change in the number of fatalities, serious injuries, slight injuries and property damage (to vehicles and fixed roadside objects).

The recommended way to summarize the results of studies is by means of a qualitative meta-analysis (ROSEBUD, 2004). The technique provides both the weighted estimate of the mean effect and a confidence interval for the estimate (a 95% confidence interval is common). The meta-analyses of evaluation studies served as a basis for “The Handbook of Road Safety Measures” (Elvik et al, 2009) where Elvik et al (2009) summarize the findings of several existing studies on roadside treatments. An example of the effects of a roadside treatment is given in Table 6.

Table 6. Effects of new guardrail along embankment in RORA (Elvik et al, 2009)

<table>
<thead>
<tr>
<th>Accident severity</th>
<th>Percentage change in the number of accidents</th>
<th>Best estimate</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatal accidents</td>
<td>-44</td>
<td>(-54, -32)</td>
<td></td>
</tr>
<tr>
<td>Injury accidents</td>
<td>-47</td>
<td>(-52, -41)</td>
<td></td>
</tr>
<tr>
<td>Unspecified</td>
<td>-7</td>
<td>(-35, +33)</td>
<td></td>
</tr>
</tbody>
</table>
3.3.2. Implementation costs of measures

The implementation costs will be determined for each of the safety measures considered. The implementation costs are the social costs of all means of production (labour and capital) that are employed to implement the measure (ETSC, 2003). The implementation costs are generally estimated on an individual basis for each intervention.

A survey of typical costs of engineering measures that can be implemented in roadside safety interventions in the context of a roadside CBA was sought in the present study.

The implementation costs should be also converted to their present values, which include both investment costs and the annual costs of operation and maintenance.

It should be emphasised that the cost figures are uncertain for a number of roadside safety measures. For example, in the case of flattening slopes there is a wide variation depending on local conditions, as regards both ground conditions and filling material availability. Also the costs of changes in cross-section (which includes clear zone widening) vary greatly, not only with the geomorphological conditions of the terrain, but also with land use (expropriation costs).

3.4. Evaluation model

The evaluation model is based on Cost Benefit Analysis (CBA), which enables decision-making and choice of the policy with the highest return in monetary terms. CBA aims to find whether a proposed intervention is economically efficient and how efficient it is (and if alterations in the intervention could make it more efficient).

The benefit-cost ratio of a road safety measure is defined as the present value of all benefits of the intervention divided by the present value of its implementation costs (CEDR, 2008):

\[
\text{Benefit } - \text{Cost ratio} = \frac{\text{present value of all benefits}}{\text{present value of implementation of all costs}} \quad (\text{eq.3})
\]

Hence, in a cost-benefit analysis, effects are compared in monetary terms. The monetary terms include accident costs, as well as a number of other factors, depending on the type and range of other effects considered, such as costs of travel time, vehicle operating costs, and external costs for impacts on air pollution and traffic noise.

The proposed procedure has one output: the recommendation of an efficient roadside safety measure for a particular road stretch.
4. **CONCLUSIONS AND FURTHER RESEARCH**

The present study will have a direct impact on the knowledge on ran-off-the-road accidents in Portugal and in the tools available to roadside design in interurban roads. These tools will improve procedures used at the design stage of new roads, at the redesign stage of existing roads and in the operational management of the road network.

Furthermore, the study will promote a faster adoption in Portugal of new technical standards, namely EN 1317 and EN 12767 (European Committee for Standardization, 2007b), which are currently not used.

An important aspect of the Project SAFESIDE is the adoption in Portugal of a procedure for in-depth analysis of ran-off-the-road accidents consistent with European recommendations, using detailed measurements and special calculation algorithms. The accident models and procedures for assessing different roadside scenarios will be developed with a view to their application by Portuguese road designers and managers. The results and tools developed in the project will be disseminated by road authorities, municipalities and relevant technical members of the road community. The tools and procedures developed will be adapted to the Portuguese road transport system, which is in agreement with the importance that subsidiarity has in road safety issues.

5. **REFERENCES**


SETRA (1999). Accidents Mortels Contre Obstacles Fixes – CEESAR.


EVALUATION OF THE SAFETY-BASED INCENTIVES IN PUBLIC PRIVATE PARTNERSHIPS (PPPs): THE CASE OF SPAIN.
1. INTRODUCTION

Many countries around the world are seeking new means to involve the private sector in managing and financing infrastructure through Public Private Partnerships (PPPs). Three reasons lay behind this trend: the growing budgetary constraints, the search for greater productivity efficiency, and the improvement of quality through a better allocation of risks and incentives (OECD, 2008). One of the most common ways of implementing PPPs in managing infrastructure is through the concession approach, which consists basically in transferring final design, construction, maintenance, and operation of the infrastructure to a private consortium, in exchange for which that consortium receives the right to charge a fee to the user or to the government on behalf of the user, for a period of time contractually agreed in advance (Vassallo and Gallego, 2005).

One of the key aspects of PPPs is to encourage the private sector to manage and operate the infrastructure in the best way. To that end, in the last few years, PPPs are evolving from mere demand-based contracts (when the revenues of the contractor are related to the traffic demand) to performance-based contracts referred to different aspects such as availability, congestion, state of the pavement, safety, and so on. Consequently, the revenues of the contractor tend to depend more and more to the quality of the services rather than on traffic demand (Harding et al., 2010). Two reasons lie behind this trend. First, PPP contractors can manage better the service performance they offer than the traffic flows in the infrastructure. And second, encouraging the PPP contractors to provide a better service by aligning the social and the private benefits will end up producing a more efficient outcome to the society.

Traditionally, most of the infrastructure management contracts, including PPPs, had not introduced explicit incentives to increase quality. However, nowadays most PPP contracts are encouraging the introduction of incentives tied to bonuses and penalties to foster the contractor to provide the optimal quality level.

Those performance-based incentives, however, have to be introduced in the right way in PPP contracts. To that end, the marginal reward to the contractor for reaching a certain quality level should never be larger than the marginal social benefit produced at that level. The contractor will provide a quality level at the point where the marginal revenue obtained due to a certain quality increase equals its marginal cost. If the incentives are defined this way, the contractor will be encouraged to provide the best service compatible with its production costs (Vassallo, 2007).

The aim of this paper is to identify whether the incentives to improve road safety in PPPs are ultimately effective in improving safety ratios with an empirical analysis for the case of Spain. The results show that there are more fatalities, injuries and accidents on road segments without incentives when they are compared with other segments with incentives.
2. **Analysis of Experiences in the Introduction of Road Safety Incentives in Europe**

Europe is the world’s region with the greater tradition of incorporating performance-based incentives in PPPs. There are few countries with PPP roads that introduce positive incentives based on explicit road safety indicators. These countries where this was done, it became a normal practice and the latest PPP contracts continue to introduce and improve these incentives. This happens in Spain, Finland, Hungary, Norway, Portugal and United Kingdom. Italy introduced price-caps tied to road safety indicators in such a way that the concessionaire is allow to set higher tolls if the safety ratios are better. In Ireland, PPP road contracts include some road safety indicators but they do not provide bonuses. In other countries, like Denmark, the Netherlands and Belgium, there are new and sophisticated PPP contracts, but they do not have positive incentives based on explicit road safety indicators.

The design of the road safety indicators is quite heterogeneous across different countries. There are differences both in the variable adopted to measure the outcome and in the final formula employed. Most of the PPPs include number of injuries, number of fatalities or a combination of number of light accidents, serious accidents and fatal accidents to build the indicator.

On the other hand, including the exposure to the risk (expressed by traffic) is a generalized practice. Very often, the initial accident data is divided by the annual traffic, usually measured in terms of millions or billions of vehicles-kilometers. E18 road (Muurla–Lohja) in Finland, the M6 road in Hungary, several PPPs in Portugal (the IP-4, for instance) and the latest PPP roads awarded in Spain use this methodology. The advantage of introducing the exposure to the risk (traffic) explicitly is that distortions in road safety results are reduced.

With the same aim, in many other PPP contracts the assessment of the indicator is done by comparing similar roads, in terms of traffic, number of carriageways, etc. With this methodology it is possible to control the global evolution of casualties by many factors, most of them not manageable by the road operator. Some road contracts where the indicator is set in this way are E-18 road (Grimstad – Kristiansand) in Norway, the latest PPP roads awarded in the United Kingdom (for example, A1 & M25) and some PPP roads in Spain (like the M-407 motorway). Regarding the way of rewarding or penalizing the contractor, it was identified two trends: incentives related to the extension of the deadline of the project, and incentives related to increase the fee to be paid to the PPP contractor.
3. Characteristics of Motorways in Spain

Spain has extensive experience in managing and financing motorways through public-private-partnerships (PPPs). Most of the PPPs have been put into effect through concession contracts that have a long tradition in Spanish administrative law. Most of the motorway concessions awarded in Spain have been toll motorways. However, in the last few years, there was a large increase in the number of other PPP approaches, such as shadow-toll\(^1\) or performance-based contracts.

Three different periods regarding the implementation of motorway concessions in Spain can be identified: from 1967 to 1975, from 1976 to 1995, and from 1996 to the present. Between 1967 and 1975, 2,042 km of toll motorways were granted by the central government of Spain. The results of the implementation of concession contracts in Spain during this period were rather controversial. On the one hand, motorway concessions achieved the goal of providing the country with a modern motorway network at a time when the public budget of Spain was not sufficient to afford such a huge cost. On the other hand, the guarantees made by the government to facilitate concessions’ funding over time became very costly for the country (Izquierdo and Vassallo, 2004)\(^2\).

The second stage of motorway concessions is from 1976 to 1995. In this period, no motorway concessions were awarded. There were several reasons for this. First, the two petroleum crises in the 1970s destabilized the Spanish economy. Second, after Franco’s death, the political atmosphere in Spain was uncertain. Third, and most important, the Socialist government, which took office in 1982 and remained until 1996, was politically opposed to promoting private concessions as a means to finance motorways. Instead, the socialist government opted

---

\(^1\) Shadow tolls involve payments per vehicle using a kilometre of the project road in accordance with a pricing structure. Different payments are due for traffic within different traffic bands and depend on the length of the vehicle (Grimsey and Lewis, 2004).

\(^2\) In a PPP transaction, very often public guarantees are provided to the concessionaire due to the risk allocation between the public and private parties. Risk allocation has a great relationship with the financial cost of the project. The greater the risk allocated to the concessionaire the higher will be the financial cost of the project. When the risk allocated to the concessionaire are very high, the financial cost of the project becomes significant, which can threaten the ultimate financial feasibility of the project. This is the reason why very often public guarantees are provided. These guarantees facilitate the financial feasibility of the project at the expense of allocating a greater risk to the government.
for modernizing the Spanish road network by widening and upgrading the most important roads, turning them into dual-roadway fast lanes. These free motorways were called “autovías”.

The term “autovías” was used in Spain in the early years of 80s for identifying any free motorway with physically separated lanes in each direction. The first autovías in Spain were built by duplicating lanes out of single carriageways. The design standards of these motorways (known as first generation autovías) had design standards well below those of the toll motorways.

Since 2000 now, the quality standards in the construction of autovías improved notably to make comparable with toll motorways. These are called second generation autovías. There is no technical difference between second generation autovías and toll motorways. The first generation autovías and second generation autovías were built, funded and managed by the Spanish government. None of the public motorways (autovías) has economic incentives to improve road safety though the government is committed to building and maintaining the autovías with socially expected quality standards.

The third stage began in 1996 and continues into the present. In 1996, the conservative Popular Party took office in Spain. The need to contain Spain’s public deficit was the most difficult challenge facing the new government. This was the main reason why the new government decided to implement once again the policy of offering concessions so as to encourage the participation of the private sector in financing new transportation infrastructure. From 1996 to now, 1,003 kilometers of new toll motorway concessions have been awarded by the central government of Spain through this approach. Another novelty of the period since 1996 is that not only the Spanish central government, but also the regional and even the local governments have started using the concession approach to implement both toll and shadow-toll motorways.

Most of the PPPs implemented in the last stage include a provision to extend the contract duration up to four years if several performance-based indicators tied to quality aspects like queuing in toll plazas, congestion, state of the pavement, safety, and satisfaction of the users are ultimately fulfilled. Other PPP’s awarded recently in Spain also include incentives in terms of annual bonuses to be incorporated to the periodic fee paid by the government to the contractor.

Regarding safety, the PPP contractor can be granted an extension of the contract if safety indicators remain below an accident benchmark for similar roads. To that propose the government measures in a yearly basis the Risk Index (RI) and Mortality Index (MI) of the motorways and compare it with other motorways with similar characteristics in terms of alignment and traffic flow.
4. Models that Explain Road Safety: A Literature Review

This paper analyzes whether the incentives to improve road safety in PPPs are effective. The statistical models most commonly used to explain the relationship between motor vehicle accidents and a set of predictor variables are the Poisson and NB regression models (Miaou and Lum, 1993; Noland and Oh, 2004; Chang, 2005; Caliendo et al., 2007).

From an empirical stand-point, the relationship between accident frequency and traffic flows can be found in Jovanis and Chang (1986), Abdel-Aty and Essam Radwan (2000) and Persaud et al. (2000); and the relationship between accident rate and traffic flow can be found in Vitaliano and Held (1991) and Hauer and Bamfo (1997). The relationship between accident frequency or accident rate and traffic flows show a great variation in their results. Jovanis and Chang (1986), Abdel-Aty and Essam Radwan (2000) and Persaud et al. (2000) point out that accident frequency increases with Average Annual Daily Traffic (AADT). On the other hand, Vitaliano and Held (1991) cannot detect any significant increase in the accident rate when AADT increases. According to Hauer and Bamfo (1997) the accident rate even decreases with an increasing AADT.

Few studies have analyzed the effect of heterogeneous flows, and specifically the effect that the presence of heavy good vehicles (HGVs) in the traffic flow has on accidents. Hiselius (2004) analyzed the relationship between accident frequency and traffic flow in four different road types according to speed limit and road width in two conditions: homogenous and heterogeneous traffic. The results show that the expected number of accidents increases less than proportionally with the traffic flow, in the homogenous case. For the heterogeneous case, the expected number of accidents decreases with increasing number of trucks. According to Arenas et al. (2009), the expected number of accident increases with the addition of one vehicle in AADT, and increases with one additional HGV when comparing high capacity roads to single carriageway roads.

Other studies have been carried out to establish relationships between accidents and the frequency of intersections (Ivan and O’Mara, 1997), environmental factor (Fridstrom et al., 1995; Shankar et al., 1995; Chang, 2005; Caliendo et al., 2007), geometric infrastructure characteristics (Hauer, 2004; Chang, 2005), number of lanes (Milton and Mannering, 1998; Noland and Oh, 2004), and speed limits (Fridstrom et al., 1995; Ossiander and Cummings, 2002). In this paper, Poisson and negative binomial (NB) regression models were applied to know the relationship between safety incentives and road safety in Spain. This paper focuses on variables related to traffic flow, infrastructure characteristics (intersections), road operation and incentives given to the PPP contractors. The author has not found any empirical study about safety incentives offered to the PPP contractors.
5. DATA FOR THE CASE OF SPAIN

The models that were calibrated in this paper cover the year 2006. We chose this year because it is the most recent year where a complete database is available. The data used for the empirical model came from two different sources: police-reported accident data supplied by the Ministry of Internal Affairs (Ministerio del Interior, 2006) and traffic data supplied by the Ministry of Public Works (Ministerio de Fomento, 2006). Using these databases it was necessary to build the final database combining both accident data and traffic data. The population for the models were made up of road stretches of the Spanish high-capacity network both PPP motorways (toll motorways) and public motorways (autovías).

Some exogenous variables that may potentially influence safety and do not depend on the concessionaire’s ability to manage the road were selected. These are: continuous variables such as (1) Average annual daily traffic (AADT), (2) Percentage of heavy goods vehicles (%HGV) and (3) Number of intersections for each stretch (INT) and two selected variables that may explain any relation between PPPs and accidents, (4) Road operation (RO) and (5) Incentives (INC). Furthermore, in order to fit the model it was considered the vehicle exposure (vk). It was measured in millions of vehicle-kilometers as $v_k = 365 l_j AADT_j/10^6$, where AADT$_j$ and $l_j$ are respectively the average annual daily traffic and length (km) of road section $j$ obtained from traffic database.

The number of intersections for each stretch (INT) introduced as discrete value with 2 levels (it takes 1 for stretches with at least 1 intersection and it takes 0 otherwise) was obtained from the Geographic Information System (GIS), the analyses was performed using ArcGis 9.2. Each stretch of the sample was analyzed using the GIS application to count the number of intersections.

Data for Road operation (RO), introduced in models as categorical variable, was obtained from the Traffic map 2006 (Ministry of Public Works). The roadway segmentation is defined by the Ministry of Public Works (Ministerio de Fomento, 2006b), which can be for the case of the Spanish high capacity network: second generation autovías (AV), first generation autovías (1AV) or toll motorways (TM). TM are PPPs, AV and 1AV are public motorways.

To know which segments had road safety incentives (INC) it was analyzed each concession contract in force in 2006. The first Spanish concession contract awarded with implementation of road safety incentives was in 2002. All of the contracts from 2002 to 2006 were analyzed. INC is a discrete variable with 2 levels (it takes 1 for stretches with safety incentives and it takes 0 otherwise).

For this study, the dependent variables are the number of fatalities, injuries and accidents. 1,042 road segments were extracted out of a total of 6,293 from the 2006 traffic map, after selection criteria based on complete information for traffic flow and infrastructure variables. It was considered stretches with and without accidents to avoid selection bias. The study
includes PPPs (toll motorways) and public motorways (first and second generation *autovías*). For this study was considered only stretches of the Spanish Interurban Road State Network. The final data base represents 62% of the total Spanish high-capacity network. The total number of segments (with and without accidents) and the length by road operation are presented in Table 1.

**Table 1. Road segment data**

<table>
<thead>
<tr>
<th>Road operation</th>
<th>Segments</th>
<th>Length (Km)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nº</td>
<td></td>
</tr>
<tr>
<td></td>
<td>with accidents</td>
<td>without accidents</td>
</tr>
<tr>
<td>AV</td>
<td>275</td>
<td>287</td>
</tr>
<tr>
<td>1AV</td>
<td>43</td>
<td>266</td>
</tr>
<tr>
<td>TM</td>
<td>60</td>
<td>111</td>
</tr>
<tr>
<td>Total</td>
<td>378</td>
<td>664</td>
</tr>
</tbody>
</table>

The descriptive statistics are represented in Table 2. The descriptive statistics indicate that the mean traffic intensity (AADT) is higher on 1AV than AV and TM. The low traffic intensity in TM could be explained because the users pay tolls. There is more heavy good vehicles traffic in public motorways (AV and 1AV) than PPPs (TM). The range and the standard deviation in AV and 1AV are higher than TM, which indicates more heterogeneity in traffic flow.

**Table 2. Descriptive statistics by road operation and total. Year 2006**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Road operation</th>
<th>Mean</th>
<th>S.D.</th>
<th>Range</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>AADT</td>
<td>AV</td>
<td>23,686.04</td>
<td>21,875.96</td>
<td>139,040</td>
<td>1,270</td>
<td>140,310</td>
</tr>
<tr>
<td></td>
<td>1AV</td>
<td>35,735.05</td>
<td>34,262.09</td>
<td>186,365</td>
<td>5,135</td>
<td>191,500</td>
</tr>
<tr>
<td></td>
<td>TM</td>
<td>16,508.43</td>
<td>12,950</td>
<td>58,093</td>
<td>1,094</td>
<td>59,187</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>26,080.92</td>
<td>26,046.307</td>
<td>190,406</td>
<td>1,094</td>
<td>191,500</td>
</tr>
<tr>
<td>%HGV</td>
<td>AV</td>
<td>18.535</td>
<td>10.12</td>
<td>72.40</td>
<td>2.80</td>
<td>75.20</td>
</tr>
<tr>
<td></td>
<td>1AV</td>
<td>23.37</td>
<td>8.95</td>
<td>48.10</td>
<td>4.60</td>
<td>52.70</td>
</tr>
<tr>
<td></td>
<td>TM</td>
<td>10.35</td>
<td>5.39</td>
<td>32.60</td>
<td>1.50</td>
<td>34.10</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>18.63</td>
<td>10.04</td>
<td>73.70</td>
<td>1.50</td>
<td>75.20</td>
</tr>
<tr>
<td>INT</td>
<td>AV</td>
<td>2.08</td>
<td>1.21</td>
<td>8</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>1AV</td>
<td>2.38</td>
<td>1.59</td>
<td>12</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>TM</td>
<td>1.54</td>
<td>0.70</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2.08</td>
<td>1.29</td>
<td>12</td>
<td>0</td>
<td>12</td>
</tr>
</tbody>
</table>
6. **METHODOLOGY AND RESULTS**

Poisson and NB regressions were applied to determine the relationship between fatalities, injuries and accidents and traffic variables, number of intersections for each stretch, road operation and road safety incentives offered to the concessionaire. In this work the response variables are number of fatalities, injuries and accidents. The Poisson overdispersion parameter indicates that the data may be overdispersed. When Pearson's chi-square divided by the degrees of freedom is greater than 1, the data may be overdispersed, otherwise the data may be underdispersed. All Poisson models presented Value/df greater than 1, therefore it was necessary to test the NB as an alternative model, which enables the variance of the dependent variable to differ from its mean.

The general equation in a segment $j$, road operation $\left(RO_{ij} = \frac{1}{3}\right)$ being $RO_3 = AV$ the reference class, and the incentives systems to the concessionaire $\left(INC_{ij} = 1 = yes \quad 0 = no\right)$ is:

$$E[Y_{ij}] = \exp[\beta_0 + \beta_1 \log(AADT_{ij}) + \beta_2 \log(\%HGV_{ij}) + \beta_3(INT_{ij}) + \alpha_i(RO_i) + \kappa_l(INC_l) + \log(vk_{ij})]$$

Where:
- $E[Y_{ij}]$ is the expected number of fatalities, injuries or accidents
- $AADT_{ij}$ is the Average Annual Daily Traffic
- $\%HGV_{ij}$ is the Percentage of heavy goods vehicles
- $INT_{ij}$ is the number of intersections for each stretch
- $RO_i$ is Road operation which could be toll motorways, first generation *autovías* and second generation *autovías*.
- $INC_i$ reflects the introduction of road safety incentives in the stretch analysed
- $vk_{ij}$ is the vehicle exposure measured in million of vehicle-kilometers.

Table 3 summarizes the estimated NB regressions. Each column refers to a model with the endogenous variables and every row to exogenous variables. There is an estimated parameters for each variable. The significance of coefficients was checked using Wald statistic (in bracket), which rejects the null hypothesis that the coefficient is zero with a level of 95% confidence. Different goodness-of-fit statistics were used to select the model such as deviance, log-likelihood and Pearson chi-square statistics. Other measures were also evaluated such as Akaike Information Criterion (AIC) (Akaike, 1974) and the Bayesian Information Criterion (BIC) (Schwarz, 1978).

A criterion for variables inclusion was used by testing the likelihood ratio (LR). In addition, the correlation coefficient between them was prevented.
### TABLE 3. Negative binomial regression models for fatalities, injuries and accidents

<table>
<thead>
<tr>
<th>Variables</th>
<th>Measurement level</th>
<th>FATALITIES</th>
<th>INJURIES</th>
<th>ACCIDENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td>0.936</td>
<td>0.334</td>
<td>-0.677</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.403)</td>
<td>(0.269)</td>
<td>(0.947)</td>
</tr>
<tr>
<td>Log(AADT)</td>
<td>S</td>
<td>-0.700a</td>
<td>-0.157a</td>
<td>-0.1068</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(37.977)</td>
<td>(10.016)</td>
<td>(4.031)</td>
</tr>
<tr>
<td>Log(%HGV)</td>
<td>S</td>
<td>-0.084</td>
<td>-0.470a</td>
<td>-0.507a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.236)</td>
<td>(37.863)</td>
<td>(37.781)</td>
</tr>
<tr>
<td>INC</td>
<td>C</td>
<td>-0.818c</td>
<td>-0.676c</td>
<td>-0.943b</td>
</tr>
<tr>
<td></td>
<td>stretch with incentives</td>
<td>(0.552)</td>
<td>(2.670)</td>
<td>(3.946)</td>
</tr>
<tr>
<td>INT</td>
<td>S</td>
<td>0.817</td>
<td>0.125</td>
<td>0.167</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.360)</td>
<td>(0.629)</td>
<td>(0.880)</td>
</tr>
<tr>
<td>ROAD OPERATION</td>
<td>C</td>
<td>-0.461</td>
<td>-0.398a</td>
<td>-0.351a</td>
</tr>
<tr>
<td></td>
<td>TM</td>
<td>(2.954)</td>
<td>(11.809)</td>
<td>(7.562)</td>
</tr>
<tr>
<td></td>
<td>1AV</td>
<td>-0.511b</td>
<td>-0.405a</td>
<td>-0.423a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(6.545)</td>
<td>(23.900)</td>
<td>(22.610)</td>
</tr>
<tr>
<td>Log(vk)</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure level: Scale (S), Categorical (C)</th>
</tr>
</thead>
</table>

*a p<0.01  
*b p<0.05  
*c p<0.10

After analyzing Tables 3, it could highlight the following results:

1. Log(AADT) is statistically significant for the fatality, injury and accident rates. The coefficient signs are negative for all models (fatalities, injuries and accidents), suggesting that a greater AADT is associated with lower rates for fatalities, injuries and accidents. This result confirms the hypothesis proposed by Vitaliano and Held (1991) and Hauer and Bamfo (1997) that accident rates decrease when AADT increases.
2. Log(\%HGV) was found to be statistically significant for injury and accident models. The coefficient signs are negative for all models. This means that the larger the percentage of heavy good vehicles the smaller the injury and accident rates. This could be related to the speed reduction that heavy vehicles impose to light vehicles on the traffic flow. This result is in line with Hiselius (2004), if this analysis was in terms of number of accidents.

3. INC variable showed the expected negative sign, suggesting that establishing incentives in the PPP contracts is associated with lower rates for fatalities, injuries and accidents. The coefficient signs do not change in all models. This variable was found to be statistically significant for injury and accident models.

4. The INT (intersection) variable showed the expected positive signs suggesting that the increased number of intersections is associated with the increased level of fatalities, injuries and accidents. There are more fatalities, injuries and accidents in stretches with at least one intersection than stretches without intersections. However, this variable was found to be statistically insignificant.

   The main reason for the insignificance of INT in the models could be that the kind of roads used for the analysis does not have at grade intersections which are those that most affect road safety. The Spanish high capacity network only has grade-separated intersections. These types of intersections reduce the accident potential due to vertical separation of traffic.

   This result is in line with the findings of Ivan and O’Mara (1997).

5. RO indicates that there is a differentiated behavior among toll motorways, second generation autovías and first generation autovías.

   Toll motorways (TM) showed the expected negative sign in all models suggesting that there are fewer fatalities, injuries and accidents in toll motorways compared to AV, which is the reference class. In other words, toll motorways managed and operated by the private sector turn out to be safer that free motorways managed and operated by the public sector.

   The only odd result that was found in the analysis is that the variable 1AV has an unexpected negative sign suggesting that 1AV has fewer fatalities, injuries and accidents compared to AV. This result is strange because, as mentioned earlier in this paper, first generation autovías (1AV) have poorer design standards than second generation autovías (AV).

   The main reason for that is that first generation autovías are among the busiest motorways in Spain in terms of traffic so the AADT in these motorways is often close to their capacity and consequently the speed of the flow is lower, which improves safety ratios.
The expressions of fatality, injury and accident rates for different road operations can be determined by the parameters estimated in Table 3. As one example, the equations of toll motorways by segment \((j)\) in specific scenarios are calculated in the following way:

\[
\hat{\lambda}_{ij} = \frac{E[Y_{ij}]}{vk_j}
\]

The average accident rates per 1 million vehicle-kilometers are:

**Toll motorways segments with incentives – without intersection:**

\[
\hat{\lambda}_{TM_j} = 0.139(\text{AADT})^{-0.106}(\%HGV)^{-0.507} = 0.015
\]

**Toll motorways segments with incentives – at least 1 intersection:**

\[
\hat{\lambda}_{TM_j} = 0.164(\text{AADT})^{-0.106}(\%HGV)^{-0.507} = 0.018
\]

**Toll motorways segments without incentives – without intersection:**

\[
\hat{\lambda}_{TM_j} = 0.357(\text{AADT})^{-0.106}(\%HGV)^{-0.507} = 0.040
\]

**Toll motorways segments without incentives – at least 1 intersection:**

\[
\hat{\lambda}_{TM_j} = 0.423(\text{AADT})^{-0.106}(\%HGV)^{-0.507} = 0.046
\]

From these expressions it can observe clearly that the incentives given to the concessionaires have a positive impact on reducing accident rates. For stretches without intersection, the accident rate per 1 million vehicle-kilometers is about 2.66 times higher in toll motorways without incentives compared to toll motorways with incentives. And, for stretches with at least 1 intersection, the accident rate per 1 million vehicle-kilometers is about 2.55 times higher in toll motorways without incentives compared to toll motorways with incentives.
7. CONCLUSIONS AND FUTURE RESEARCH

The main conclusions of this paper are the following:

– Some variables such as the AADT and the %HGV, which are not manageable by the contractor, are the most significant predictors of fatalities, injuries and accidents.

– The INC variable was found to be statistically significant in the injury and accident model. However, if incentives have an impact on the frequency of road accidents, they have an impact on fatalities as well, because of the negative sign in fatal models. If incentives given to the concessionaires avoid accidents, they avoid fatalities which are a consequence of accidents. The results indicate that incentives have an influence on road safety, there are more fatalities, injuries and accidents on road segments without incentives than in road segments with incentives. This fact demonstrates that the implementation of road safety incentives in PPP contracts is effective to encourage the contractor to adopt the measures that he can handles to improve road safety standards.

– The results indicate the need for further research. First, the analysis of the size of the economic incentive set up in the PPP contract on the ultimate improvement of safety ratios is a crucial aspect. Second, it would be useful to analyze the evolution of safety performance over the years in PPP motorways with safety-based incentives. Third, a cross-cutting comparison of the different types of incentives (incentives related to the deadline of the project vs incentives related to payments) would be of the greatest interest.
REFERENCES


Universidad Politécnica de Madrid. Ministerio de Fomento.


EFFECT OF TYRE TREAD DEPTH ON ACCIDENT INVOLVEMENT DURING SUMMER IN FINLAND

ABSTRACT

The aim of the paper is to determine whether various tyre tread depths in passenger cars and vans were associated with different fatal accident involvement rates or different weather distribution of fatal accidents. Data used in this study includes fatal accidents in 2000–2007 reported by Finnish road accident investigation teams, and tyre check data from annual tyre safety campaigns.

The risk of being responsible for a fatal crash for drivers having an illegal tread depth of less than 1.6 mm is three times as high as for those having a tread depth of 3.5 mm or above. Among drivers less responsible for a crash, the risk of the being involved in a fatal crash was similar or slightly lower with a low tread depth than with a higher one. However, there are behavioural factors that correlate with the tyre tread depth. Accident involvement risk in relation to tread depth will be compared with compliance with the main traffic rules (e.g. use of seatbelts, exceeding the posted speed limits and driving while intoxicated). Among drivers who complied with the main traffic rules, those with a tread depth below 1.6 mm were just 20% more likely to be involved in fatal crashes than those with a depth of 3.5 mm or more. Consequently, the main results imply that drivers should be informed about the importance of tyre depth as a part of their general driver training, even if the increase of risk due to low tyre tread depth is not as high as often considered.

INTRODUCTION

It has long been known that a decrease in a vehicle’s tyre tread depth leads to increased stopping distance and a greater likelihood of aquaplaning on wet asphalt. But is it possible to see this effect in accident statistics?

Relatively few studies have evaluated the effect of tyre tread depth on reported accidents. Hankins et al. (1971) studied vehicle and pavement factors in wet-weather accidents in Texas.
Accident vehicles had less tyre tread depth compared to the average vehicle in the study area. For skidding accidents in wet weather, tread depth was the second most important out of the studied five factors (road surface texture, tread depth, friction, speed, tyre pressure). Dijks (1976) cites a report by the Highway Safety Foundation (1971), which found that involvement in accidents increases with decreasing tyre tread depth on dry and wet road surfaces alike. For tyre tread depths less than 0.8 mm, accident involvement is far more common on wet roads than on dry roads. Good et al. (1987) studied collisions with utility poles in Australia and compared tyre tread depths of collision vehicles with a random sample of vehicles on the road. They noticed that accident involvement increased markedly for tread depths less than 3 mm, particularly on wet roads. Vehicle A vehicle with a tread depth of only 0.5 mm was about 15 times more likely to be involved in an accident than a vehicle with 5 mm tread depth. Based on the results, Good et al. concluded that on a safety basis, the minimum legal tread depth for tyres should be 3 mm. Fosser and Ingebritsen (1991) studied the effect of tyre condition on winter accidents. The results show that tyres with high friction (rich studding or great tread depth) increased safety on snowy and icy roads. The reduction in tread depth of from 7 mm to 5 mm increased the likelihood of an accident by 8.1%. Tyre condition did not correlate with accident risk on dry and wet roads. Results The results of this study indicated that the safety effect of fit tyres would be even higher greater if drivers did not compensate it with higher speeds. Elvik et al. (2009) refer to four studies concerning tyre tread depths effect on accidents in "The handbook of road safety measures". Based on two of these studies (Glad 1988, Hankins et al. 1971), Elvik et al. estimate that increasing tread depth from less than 2 mm to 2–3 mm cuts accidents by 19%. Increasing tread depth from 2–3 mm to 3–5mm cuts accidents by 9%.

In Finland, road accident investigation teams consider the tyres as a factor contributing to the accident in every seventh fatal car accident. It is the most common vehicle-related contributing factor in fatal accidents in Finland. (VALT 2009)

In Finland, winter tyres are compulsory from 1st December to the end of February (Ministry of Transport 1992). Studded tyres may be used from 1st November until the first Monday after Easter. Most cars are equipped with two sets of tyres: a set of winter tyres (often studded) and a set of summer tyres. The minimum legal tread depth for summer tyres is 1.6 mm in Finland. Autonrengasliitto ry (Tyre Specialists of Finland, an organisation of tyre companies) recommends a minimum tread depth of 4 mm for summer tyres based on friction and aquaplaning measurements.

Autonrengasliitto organises a tyre safety campaign in Finland nearly every autumn together with the police and Liikenneturva (Central Organization for Traffic Safety in Finland). It is a nationwide campaign consisting of an active supply of information for the media and tyre checks on the roads. This campaign has been organized since 1997 in August/September with the exception of 2001 and 2006. Tyre checks take place all over the country; for example, in
autumn 2007 they covered 85 municipalities. Between 11 000 and 15 000 cars or vans are checked annually. Drivers with illegal tyres do not get a fine, but inspectors distribute a brochure entitleds “Check your tyres in time”. Lahti and Savolainen (2010) have calculated based on this tyre check data that the share of cars having at least 4 mm tyre tread depth has increased from 54.5% for 1997 to 69.0% for 2009. Share The share of cars with 0–2 mm tyre tread depth has decreased accordingly fallen from 24.4% to 13.0%, consequently.

The aim of this paper is to determine whether the various tyre tread depths in passenger cars and vans were associated with different fatal accident involvement rates or different weather distribution of fatal accidents.

**METHOD**

In this study, two methods were used to estimate the effect of tyre tread depth on accident involvement: calculation of fatal accident involvement rate per kilometres driven with different tread depths, and calculation of accident distribution on dry and wet road surfaces with different tread depths (Figure 1). The study was restricted in to summer period (May–October) accidents because of the lack of wintertime tyre check data.

![Figure 1: Method of the study.](image-url)
Autonrengasliitto tyre check data and vehicle kilometrage statistics were used for estimation of kilometrage driven with different tread depths. Tyre check data contained included the number of checked cars and vans by municipality and by minimum tyre tread depth. Minimum tyre tread depths were classified in four classes: less than 1.6 mm (illegal), 1.6–2 mm, 3 mm, 4 mm or more. Based on this data, tyre tread depth distribution was calculated by county, and it was weighted by the county’s share of vehicle kilometrage to get nationwide kilometrage distribution for tyre tread depths. The summer period’s share of annual vehicle kilometrage was estimated to be 55% based on statistics received from the Finnish National Road Administration.

The number of cars and vans involved in fatal accidents with different tyre tread depths was calculated from the database of road accident investigation teams. The fatal accident involvement rate per 100 million kilometres driven was calculated with this accident data and kilometres driven with different tyre tread depths.

Information on road surface weather was also taken from the fatal accident database of road accident investigation teams’ fatal accident database. Based on weather information, fatal accidents were classified by weather into three classes: dry road surface, wet road surface, and snowy or icy road surface (rare in the summer period). Small tyre tread depth increases the risk of aquaplaning and extends the braking distance on wet asphalt. On dry asphalt, smoothness of the tyres does not reduce friction. Thus, if worn-out tyres would were to increase the fatal accident risk, a higher than average proportion of accidents with those such tyres should be in wet weather.

A car’s tyre tread depth may correlate with the driver’s behaviour. For example, a driver with worn-out tyres may also take other risks, or on the other hand careful drivers may compensate for the effect of small tyre tread depth by driving even more carefully. The effect of behavioural factors was studied by dividing drivers in fatal accident data into two groups:

- Drivers who during the accident complied with the main traffic rules: having a driving licence, not intoxicated, using a seatbelt, following the speed limit or exceeding it by no more than 10 km/h.
- Drivers who during the accident broke some of the main rules: no driving licence, intoxicated, not using a seatbelt or exceeding the speed limit by more than 10 km/h.

Accident distribution on dry and wet road surfaces with different tread depths was calculated also for this these two driver groups. The fatal accident involvement rate (accident participations per 100 million kilometres driven) with different tyre tread depths could not be calculated reliably because of lack of kilometrage data of for this these two driver groups, but
it was estimated from the aforementioned all cars’ and vans’ tyre tread depth distribution of all cars and vans.

**DATA**

In Finland nearly all the fatal accidents since the early 1970s have been studied in depth by investigation teams. They consist of a police officer, road engineer, vehicle engineer, physician and occasionally a psychologist. Among other things the investigation team’s vehicle engineer measures tyre tread depths, among other things. The smallest of the tyre tread depths which is recorded in the accident database. Tyre tread depths are reported to tenth of a millimetre accuracy, but the real accuracy is probably 0.5–1 mm as the results cumulate to even figures.

In this study all investigated fatal accidents in May–October during the period 2000–2007 were included. All passenger cars and vans, totalling 1486 vehicles, were extracted from this data. According to estimations by investigations teams, of these vehicles 1027 or 69% were chiefly responsible for causing the accident. Of the 1486 vehicles 463 were in a single vehicle accident, 798 were in a multiple vehicle accident, 186 collided with a pedestrian or cyclist and 39 collided with an animal.

The Autonrengasliitto tyre check statistics by municipality were obtained for the years 2000, 2002–2005 and 2007. During a tyre check, police officers guide a representative sample of traffic flow to the checking point (all cars and vans are guided to check the checking point until the checking area is full, then the police waits until once all the vehicles have been checked, and then a new queue of vehicles is guided to the checking area point). Tyre tread depth is measured to 1 mm accuracy, but in such a way that an illegal, less than 1.6 mm tread depth is still distinguished from a 2 mm tread depth.

Tyre checks are carried out in the autumn, when summer tyres can be in their most worn condition. Summer period kilometrage may thus be driven with slightly deeper tyre treads than those measured in tyre checks. Assuming that the average life span of a summer tyre is four or five summer periods (~ 10 000 km driven per summer period), tyre tread depth reduces from 8 mm to 2–3 mm. In this case, the average tyre tread depth during the summer period is about 0.5 mm greater than those depths observed in the autumn. The difference between observed tyre tread depth and summertime average tread depth is therefore likely to be relatively small.

Annual passenger car and van kilometrage was obtained from Finland’s official vehicle kilometrage statistics (Finnish National Road Administration 2009). The county distribution
of vehicle kilometrage was extracted from LIISA calculation software (Road traffic exhaust gas emissions calculation software) (Mäkelä et al. 2008).
RESULTS

Kilometrage, accident participation and accident risk by tyre tread depth

Summer period car and van kilometrage was on average 25 655 million vehicle kilometres in 2000–2007. Of this, 4% was driven with illegal <1.6 mm tyre tread depth of less than 1.6 mm (Table 1), 10% with 2 mm minimum tyre tread depth, 19% with 3 mm tyre tread depth, and 67% with a tyre tread depth of 4 mm or more.

Table 1. Summer period car and van kilometrage by tyre tread depth, average for 2000–2007.

<table>
<thead>
<tr>
<th>Tyre tread depth, mm</th>
<th>Average kilometrage, million km</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–1.5</td>
<td>1 014</td>
<td>4%</td>
</tr>
<tr>
<td>2</td>
<td>2 635</td>
<td>10%</td>
</tr>
<tr>
<td>3</td>
<td>4 874</td>
<td>19%</td>
</tr>
<tr>
<td>4 or more</td>
<td>17 131</td>
<td>67%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>25 655</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

Eight percent of cars and vans involved in a fatal accident during the summer period had <less than 1.6 mm tyre tread depth (Table 2). Ten percent of vehicles had a tyre tread depth of 1.6–2.4 mm. Participants chiefly responsible for the an accident had on average more worn tyres than other participants: 10% of main cause participants had an illegal tyre tread depth, compared with 3% of other participants. A tyre tread depth of at least 3.5 mm was present in 55% of main cause vehicles and in 64% of other participant vehicles.

Table 2. Tyre tread depth in cars and vans involved in summer period fatal accidents, total for 2000–2007.

<table>
<thead>
<tr>
<th>Tyre tread depth, mm</th>
<th>Number of vehicles</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–1.5</td>
<td>115</td>
<td>8%</td>
</tr>
<tr>
<td>1.6–2.4</td>
<td>150</td>
<td>10%</td>
</tr>
<tr>
<td>2.5–3.4</td>
<td>239</td>
<td>16%</td>
</tr>
<tr>
<td>3.5 or more</td>
<td>856</td>
<td>58%</td>
</tr>
<tr>
<td>Unknown</td>
<td>126</td>
<td>8%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1 486</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

The risk of being chiefly responsible for a fatal accident is nearly triple (1.18/0.42 = 2.81) with illegal <less than1.6 mm tyre tread depth compared to that with a tyre tread depth of 3.5
mm (Figure 1). With a legal but small tyre tread depth of 1.6–2.4 mm the risk of being the chiefly responsible participant is 36% greater than with tyres having a tread of least 3.5 mm. The risk of being involved in a fatal accident as a less responsible participant is the same as or even slightly lower with worn tyres than with a tread depth of at least 3.5 mm.

![Figure 21: Risk among passenger cars and vans of being involved in a fatal accident from May to October by tyre tread depth.](image)

**Accident participation on dry and wet roads**

Roughly 20% of cars and vans are involved in a fatal accident in wet conditions regardless of tyre tread depth (Figure 2). The total number of other than main cause participants with <less than 1.6 mm tyre tread depth was very small, totalling only 13 vehicles; and so thus their 92 % share of accidents on dry roads may be due to statistical fluctuation.
Figure 32: Road surface distribution of fatal accidents in the summer period, by tyre tread depth.

Accident risk for two driver groups formed on the basis of compliance with the main traffic rules

When car and van drivers in fatal accident data were divided into two groups on the basis of compliance with the main traffic rules, each group consisted of 680 drivers. Drivers with missing traffic rule compliance data were excluded. Drivers who complied with the main rules were on average older than drivers who broke the rules (Table 3).

Table 3. Age distributions in studied driver groups.

<table>
<thead>
<tr>
<th>Age group</th>
<th>Drivers who broke the main rules</th>
<th>Drivers who complied with the main traffic rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>–25</td>
<td>38%</td>
<td>21%</td>
</tr>
<tr>
<td>26–35</td>
<td>21%</td>
<td>11%</td>
</tr>
<tr>
<td>36–45</td>
<td>13%</td>
<td>17%</td>
</tr>
<tr>
<td>46–55</td>
<td>13%</td>
<td>17%</td>
</tr>
<tr>
<td>56–65</td>
<td>6%</td>
<td>14%</td>
</tr>
<tr>
<td>66–</td>
<td>8%</td>
<td>20%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

When a fatal accident occurred, drivers who broke the rules had on average more worn-out tyres than drivers who complied with the rules. Their tyres were also more worn out than on average in cars and vans driven on Finnish roads (Figure 3). 70% ofOf vehicles with illegal
less than 1.6 mm tyre tread depth 70% had a rule-breaking driver, when while the corresponding figure for at least 3.5 mm tyre tread depth is was 44%.

The distribution of tyre tread depth of among vehicles manned by rule-breaking drivers’ vehicles is highly significantly different from the observed distribution in tyre checks ($\chi^2 (3) = 185.1, p = 0.00$). Drivers who complied with the main rules had an almost significantly different tyre tread depth distribution compared with the observed distribution in tyre checks ($\chi^2 (3) = 8.5, p = 0.04$).

If the kilometrage of rule-breaking drivers’ vehicles kilometrage was distributed for tyre tread depth similarly to that of all cars and vans based on tyre checks, their risk of being involved in a fatal accident with illegal <less than 1.6 mm tyre tread depth would be more than triple compared with a tyre tread depth of 3.5 mm or more (Figure 4). With a tyre tread depth of 1.6–2.4 mm the risk would be 1.5-fold that with a tyre tread depth of 3.5 mm or more. Similarly, if drivers who complied with the rules had their vehicle kilometrage distributed for tyre tread depth similarly to all vehicles in tyre checks, their risk of being involved in a fatal accident with illegal <less than 1.6 mm tyre tread depth would be 1.2-fold compared with a tyre tread depth of 3.5 mm or more (Figure 5).

Note that there are no precise risks in Figures 4 and 5, only relative risks so that the risk for a tyre tread depth of 3.5 mm or more is 1. Total vehicle kilometrage for rule-breaking and rule-complying driving is unknown.
If drivers breaking the main traffic rules drive a bigger share of their kilometrage with worn-out tyres than drivers on average, the relative risk with illegal tyre tread depth is less than 3.5-fold compared with the risk with at least 3.5 mm tyre tread depth. Similarly, if drivers
who comply with the main traffic rules drive a bigger share of their kilometrage with good tyres, the relative risk of illegal tyre tread depth is more than 1.2-fold compared with the risk with at least 3.5 mm tyre tread depth.

**Accident participation on dry and wet roads for two driver groups formed on the basis of on compliance with the main traffic rules**

For both driver groups 19% of fatal accidents occurred during other than dry weather. Drivers who broke the main traffic rules and had an illegal tread depth (<1.6 mm) were less likely than average to be involved in a fatal accident on other than a dry road surface (9% of participations) (Figure 6). Drivers who complied with the main rules and had an illegal tread depth were more likely than average to be involved in a fatal accident on other than a dry road surface (37% of participations). Also for tyre tread depths of 1.6–3.4 mm drivers who complied with the main rules were more likely than average to be involved in a fatal accident on a damp or wet road (22–23% of participations).

![Figure 76: Road surface distribution of fatal accidents in summer, by tyre tread depth and main traffic rule compliance.](image)

On average, two car or van drivers per year are in compliance with the main traffic rules when involved in a fatal accident and having an illegal tyre tread depth on a road surface that is other than dry. If there were only one such accident involvement instead of two, their fatal accident road surface distribution (Figure 6, fifth column) would be similar to that for drivers with a tyre tread depth of least 3.5 mm (Figure 6, eighth column). This corresponds to a 315% drop in fatal accident participation for rule-complying drivers with illegal tyres.
reduction, if they were to have a 3.5 mm tyre tread depth instead, and a 3 % drop in fatal accident participation for all drivers with illegal tyres, if they drivers would have to have a 3.5 mm tyre tread depth instead of less than 1.6 mm.

Similarly, drivers who complied with the rules and had a tyre tread depth of 1.6–2.5 mm should have 0.6 fewer wet road surface accidents involvements annually in order to have the same fatal accident weather distribution as drivers with a tyre tread depth of at least 3.5 mm. For a tyre tread depth of 2.5–3.4 mm the corresponding figure is 1.0 annual wet road surface accident involvements.

**CONCLUSION**

In During the summer period in Finland, a car’s tyre tread depth correlates with the driver’s risk-taking. A vehicle with illegal less than 1.6 mm tyre tread depth is 2.8 times more likely to cause a fatal accident than a vehicle with at least 3.5 mm tyre tread depth. However, this is not mainly not due to those the tyres, because the share of damp or wet road accidents was roughly the same, at 20%, for worn-out tyres and for tyres with at least 3.5 mm tread depth. 70 % ofAmong cars and vans involved in fatal accidents with illegal tyre tread depth, 70% had a driver who drove intoxicated, was speeding by more than 10 km/h, had no driving licence, or was not using wearing a seatbelt.

Does tyre tread depth have an effect on the accident risk of Finnish drivers breaking the main traffic rules (driving licence, seatbelt, driving intoxicated, speeding)? InNo such effect was detected in this study such effect could not be seen. Accident risk was highest but the share of wet or damp weather accidents lowest for illegal <1.6 mm tyre tread depth of less than 1.6 mm. Slick tyres do not reduce the friction on dry asphalt.

Does tyre tread depth have any effect on the accident risk of Finnish drivers who comply with the main traffic rules? Yes it does, if the tyre tread depth is illegal. Drivers who complied with the main traffic rules but had an illegal tyre tread depth had a higher accident risk and bigger share of wet or damp weather accidents than drivers with a tyre tread depth of least 3.5 mm. Therefore, this increase in accident risk is probably due to worn-out tyres. It was estimated in this study that if all drivers who drive sober, follow the speed limit and use wear a seat belt but have an illegal tyre tread depth, would were to have a 3.5 mm tyre tread depth instead, it would reduce their fatal accidents by 315%. This is much less than previous studies indicate, and the number of this kind of drivers in fatal accidents is rather small. This study did not provide clear evidence of a difference in fatal risk with a tyre tread depth of 1.6–3.4 mm compared to at least 3.5 mm depth. With these tyre tread depths the accident risk was the
same or even smaller than with a greater tread depth, but for drivers who complied with the main traffic rules the share of wet or damp weather accidents was slightly bigger.

Finnish tyre safety campaigns put emphasis onemphasize the fact that the tyres are being a factor contributing to the accident in every seventh fatal car accident, based on fatal accident investigations. This is true butAlthough this is true, based on this study it is, somewhat misleading. A big share of those these tyre-related fatal accidents also include also drunk driving and/or speeding as more important contributors, but the public probably understands this message so to mean that better tyres for normal rule-complying drivers would save lots of lives. The results emphasize the importance of linking tyre information with other safe driving instructions and traffic control, because the vast majority of drivers with worn-out tires in fatal accidents also take other significant risks.

This study’s weather distribution figures would have benefited from an extra aspect: paved and unpaved roads. A small share of fatal accidents in the data had occurred on unpaved roads, and worn-out tyres reduce friction on wet asphalt but also on dry loose gravel. It is possible that some small share of the high accident rate among rule-breaking drivers with small tyre tread depth on dry roads is due to the combination of those such tyres and loose gravel.

This study could be extended with an examination of summer weather distribution. Traffic kilometrage in different weathers and road surfaces including paved roads and gravel roads could be estimated combining the Road Administration’s weather station data with traffic counter data. Thus it should be possible to calculate the fatal accident risk in different types of weather.

ACKNOWLEDGMENTS

The author would like to thank Autonrengasliitto ry for tyre check data and the Finnish Motor Insurers’ Centre for accident data. This study was implemented as part of the research programme “Safe traffic 2025” financed by VTT Technical Research Centre of Finland, A-Katsastus Group, the Finnish Rail Administration, the Finnish Rail Agency, the Finnish Road Administration, Michelin Nordic AB, the Ministry of Transport and Communications Finland, Neste Oil Corporation and VR-Group.

REFERENCES


RECOGNISING TRANSITIONS BETWEEN RURAL ROAD CATEGORIES: THE ROLE OF ROAD LAYOUT AND INTERSECTION TYPE

Agnieszka Stelling, SWOV Institute for Road Safety Research, The Netherlands
agnieszka.stelling@swov.nl

ABSTRACT

Moving from one road to another and traversing different road categories requires drivers to adjust their expectations concerning speed limit and road users to be encountered. Recent research shows that recognisable road layout can support drivers in forming correct expectations about transitions between road categories. The role of the road layout in combination with the type of intersection has not yet been investigated. This study aimed to identify the optimal conditions under which drivers recognise transitions from one road category to another. The experiment was performed with a series of animated video scenarios showing footage of the imaginary driver driving along a road, arriving at an intersection, turning right and then driving along a road of a different category. Participants \( N = 160 \) had to indicate in a web-based survey their expectations regarding speed limit and access restriction of the road before and after a transition. The scenarios varied on three aspects: (1) type of transition: transitions between distributor and through roads (DR-TR) and transitions between distributor and access roads (DR-AR); (2) type of intersection: typical (common for a particular type of transition) and less typical (which can be found with various types of transition); (3) type of road layout: distinguishable and less distinguishable in terms of whether the layout of the road before differs much from the layout of the road after a transition. Half of the participants watched animation clips showing transition from a lower to a higher order road, i.e. DR-TR and AR-DR transitions, and the other half from a higher to a lower order road, i.e. TR-DR and DR-AR transitions. The study indicates that typical intersection types and distinctive layout of road sections can enhance recognisability of transitions, especially DR-TR transitions. The results are discussed in the context of road safety.

Keywords: recognisability, change blindness, road design, expectations
1. INTRODUCTION

Forming correct expectations while driving is very important as they help drivers to quickly focus attention on the most relevant aspects of the situation and to anticipate the most probable dangers (van der Hulst et al., 1999). Research shows that expectations are crucial in collision avoidance (van der Hulst et al., 1999). In fact a substantial number of crashes is a result of inappropriate expectations (e.g. Malaterre, 1990). One way to assist drivers in forming correct expectations about the road they are driving on is by predictable and recognisable road layout. For example, Dutch research shows that red cycle lanes evokes correct expectations about the possible presence of moped riders and cyclists (Kaptein and Theeuwes, 1996) while emergency lanes, guardrails and gantries are associated with high speeds (Theeuwes, 1994). A recognisable road layout is meant to provide information about the road type, types of road users allowed on the road and driving behaviour expected of road users. By means of evoking correct expectations, a recognisable road layout can make the traffic behaviour more predictable, prevent indecisive behaviour of road users and allow them to act more on routine. As routine behaviour is related to less serious errors types (Reason, 1990) a recognisable road layout can in the end prevent errors that could lead to crashes (e.g. Wegman et al., 2008).

1.1. Recognizable road layout

The importance of predictable and recognisable road environment has been emphasized in the Dutch safe system approach Sustainable Safety (Wegman et al., 2008). Predictability constitutes one of the sustainable safe principles and it postulates that the road features should tell immediately what road type one is driving and what is expected of each type of road users (Wegman, et al., 2008). The principle of predictability builds on two other original principles: functionality and homogeneity (see Table 1). First of all, roads should have but one function: either enabling traffic to flow or giving access to destinations (the functionality principle) and their layout need to facilitate homogeneous use in speed, mass, and direction (homogeneity).

<table>
<thead>
<tr>
<th>Sustainable Safety Principle</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functionality of roads</td>
<td>Monofunctionality of roads as either through roads, distributor roads, or access roads in a hierarchically structured road network</td>
</tr>
<tr>
<td>Homogeneity of mass and/or speed and direction</td>
<td>Equality of speed, direction, and mass at moderate and high speeds</td>
</tr>
<tr>
<td>Predictability of road course and road user behaviour by a recognizable road design</td>
<td>Road environment and road user behaviour that support road user expectations through consistency and continuity of road design</td>
</tr>
<tr>
<td>Forgiveness of the environment and of road users</td>
<td>Injury limitation through a forgiving road environment (physical) and anticipation of road user behaviour (social)</td>
</tr>
<tr>
<td>State awareness by the road user</td>
<td>Ability to assess one's capacity to handle the driving task</td>
</tr>
</tbody>
</table>
The original Sustainable Safety principles (functionality, homogeneity and predictability) together with physical forgivingness aim to create a traffic system tailored to human characteristics. The two other principles, introduced later on, social forgivingness and state awareness, focus on the role played by the road users themselves in preventing crashes and/or minimizing injury (Wegman et al., 2008).

Sustainable Safety approach distinguishes three road categories: (1) through roads (TRs): high speeds (100 or 120 km/h), physical separation of driving directions, access restrictions for mopeds, bicyclists, and agricultural vehicles, (2) distributor roads (DRs): intermediate speeds on road sections (80 km/h) and low speeds at intersections; physical separation of slow/vulnerable road users and fast traffic is recommended (access restriction at road sections for mopeds, bicyclists, and agricultural vehicles), and (3) access roads (ARs): all traffic allowed which requires low speeds (60 km/h). Ideally, recognisability and predictability is an integral characteristic of the functionality and homogeneity principles, which can be found in a corresponding layout. A starting point to make Dutch roads recognisable and predictable has been defining the so-called ‘Essential Recognisability Characteristics’ (ERCs) (CROW, 2004). These ERCs consist of road characteristics that are continuously visible to drivers, mainly variations in edge marking and the type of separation of driving directions in unique combinations of patterns per road category. Table 2 presents a number of examples of rural road layouts with ERCs for the three road types.

<table>
<thead>
<tr>
<th>Road type</th>
<th>Variants of rural road layout with Essential Recognisibility Characteristics (ERCs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Through roads (TRs)</td>
<td>1 &amp; 7 single carriageway with a broken centre line marking; 2 &amp; 8 single carriageway with a continuous centre line marking; 3 single carriageway with a broken centre line marking with green; 4 single carriageway with a continuous centre line marking with green; 5 &amp; 9 single carriageway with a curb; 6 &amp; 10 dual carriageway with a central reservation; 11 brick road without road marking; 12 asphalt road without marking; 13 asphalt road with marking to the edge; 14 asphalt road with marking further away from the edge.</td>
</tr>
<tr>
<td>Distributor roads (DRs)</td>
<td>7 &amp; 15 single carriageway with a broken centre line marking; 8 &amp; 10 single carriageway with a continuous centre line marking with green; 9 &amp; 11 single carriageway with a curb; 12 asphalt road without marking; 13 asphalt road with marking to the edge; 14 asphalt road with marking further away from the edge.</td>
</tr>
<tr>
<td>Access roads (ARs)</td>
<td>11 single carriageway with a broken centre line marking; 12 &amp; 13 single carriageway with a continuous centre line marking with green; 14 &amp; 15 single carriageway with a curb; 16 &amp; 17 dual carriageway with a central reservation; 18 brick road without road marking; 19 asphalt road without marking; 20 asphalt road with marking to the edge; 21 asphalt road with marking further away from the edge.</td>
</tr>
</tbody>
</table>
It has to be noted that drivers (can) rely on far more than the defined ERCs, such as road width and surface, curvature, and road environment characteristics (e.g. Davidse et al., 2004; Goldenbeld and van Schagen, 2007a; Kaptein and Claessens, 1998; Weller et al., 2008) when recognising a road category. Furthermore, recognition of categories is enhanced when the differences between categories are sufficiently large and the variation within each category is small (Kaptein and Claessens, 1998; Theeuwes and Godthelp, 1995). As can be seen from Table 2 it is however not always the case for the roads equipped with ERCs. Roads of different categories, i.e. TRs and DRs, look very similar. The only distinction on bases of which a distinction between TRs and DRs can be made is a difference in the edge marking (continuous: TRs versus broken: DRs). Previous research shows however that this change in edge marking is hardly noticed by road users (Stelling-Konczak et al., 2011). A more effective distinctive characteristic of a road category change is the separation of the driving direction. When the separation of the driving direction is identical (both a TR and a DR has a centre line marking) it is difficult to notice a change in a road category. Table 2 also shows that roads of the same category are not very much alike. Given the current variation within the layout of road categories in the Netherlands, and the relative small distinction in ERCs between the layouts of the different categories, it is not surprising that the expectations of road users regarding speed limits, manoeuvres and road user types allowed on the road are not always correct (e.g. Aarts and Davidse, 2008). Furthermore, recognisability of roads can be also undermined by inconsistent implementation of speed limit and access restrictions within a road category. For example, on a number of DRs in the Netherlands agricultural traffic is allowed, due to lack of alternative route options. Access roads (ARs) with speed limit of 80 km/h (instead of 60km/h) are also common, particularly when separate bicycle paths are available.

1.2. Recognisability of transitions

Moving from one road category to another requires drivers to adjust their expectations concerning speed limit and the type of road user to be encountered on the road. The question is however, whether drivers will timely detect a road category change. Fundamental psychological research shows that people often fail to detect changes in the visual field (e.g. Rensink et al., 1997; Simons and Levin, 1997). This phenomenon, called change blindness can also take place during driving (e.g. Martens, 2007; Velichkovsky, 2002). Since road category changes often take place in the vicinity of intersections, where the driver has to process a lot of information, overlooking a transition is a real possibility. Drivers can apparently be assisted in detecting a road category change by a recognisable design of a road category before and after a transition. Recent study shows that the adjustment of expectations required when moving from one road category to another can be enhanced by a recognisable and predictable layout of the road categories themselves (Stelling-Konczak et al., 2011). To start with, transitions between a DR and an AR are better recognised when no markings on
AR are present. For transitions between a DR and a TR the physical separation of driving direction (e.g. central reservation of concrete curb) turns out to be a better distinctive characteristic than the currently used edge marking. Besides the physical separation, the green marking on TR turns out to enhance recognisability of this type of transition. However, Stelling-Konczak et al. (2011) also demonstrate that in many cases the design of the road category itself is not to be sufficient enough to signalise a change of road category. Drivers make many mistakes, especially when presented with transitions between a TR and a DR. The authors conclude that when both a road before and after a transition is equipped with a centre marking (as it holds to a TR and a DR) change blindness can occur. Since road layouts are not always distinguishable enough, additional cues may be needed to draw extra attention to the category change. Therefore Stelling-Konczak et al. (2011) suggest that making a transition itself extra salient by the specific design of the intersection can be of benefit to recognisability.

1.3. Present study

The study of Stelling et al. (2011) has focused on the influence of predictable layout of the road categories themselves. The present study explores the recognisability of rural roads in transitions from one road category to another by investigating the role of road layout in combination with the type of intersection. The objective was to identify the optimal conditions under which drivers recognise transitions from one road category to another. The experiment was performed with a series of animation clips showing footage of the imaginary driver driving along a road, arriving at an intersection, turning right and then driving along a road of a different category. The clips showed two types of transitions: transitions between a TR and a DR and transitions between a DR and an AR comprising either distinguishable or less distinguishable road layout and either typical or less typical type of intersection (see also Table 2).

Distinguishable and less distinguishable road layout relates to whether the layout of the road before differs much (distinguishable) or little (less distinguishable) from the layout of the road after a transition. This distinction was based on the findings of the above mentioned study of Stelling-Konczak et al. (2011). By typical intersection type is meant a sort of intersection that is common for a particular type of transition, as opposed to a less typical one, which can be found with various types of transitions. The typical and less typical intersection types were chosen in consultation with Dutch traffic engineers. The following hypotheses were tested:

1. A transition between two road categories is better recognized when the individual roads are easily distinguishable from each other (regardless of the type of intersection).
2. A transition between two road categories is better recognized when it comprises an intersection typical for this particular type of transition (regardless of the layout of the road categories).
3. A transition comprising distinguishable road sections and a typical intersection type is better recognised than a transitions consisting of less distinguishable road sections and less typical intersection type.

4. Easily distinguishable road sections and less typical intersection type making up a transition are about as well recognized as intersections consisting of less well distinguishable road sections and a typical intersection.

2. METHODS

2.1. Participants

A stratified sample of 231 Dutch car drivers (116 males) was used in the experiment to reflect the main driving population with regard to age. The participants were panel community members and they filled out a survey on the internet. They were aged 18–70 years (M= 43 years). Almost 70% of the participants drove less than 10 000 km a year and 13% drove more than 20.000 km a year (while the Dutch drivers cover annually on average 14 000 km).

2.2. Material and procedure

The test material used in the study consisted of 16 short animation clips taken from the perspective of a car driver who drives along a road of one category, turns right at an intersection and follows a road of another category. The participants were asked to imagine they were the driver. The animation clips consisted of two sets: 1) eight animation clips presented transitions from a higher to a lower order road category i.e. TR-DR and DR-AR transitions and 2) the other eight showed transitions from a lower to a higher road category i.e. DR-TR and AR-DR transitions (see also Table 3). To control for effects of the presentation order, half of the participants watched clips showing transition belonging to the first set (TR-DR and DR-AR transition) and the other half watched clips belonging to the second set (DR-TR and AR-DR transitions). Participants were randomly assigned to one of the two sets: each person watched 8 animation clips. The clips within each set were presented at random. After each clip participants had to indicate their expectations about speed limit and access restrictions of the road before and after the intersection (see Fig.2). Responses of expectations were given by ticking the boxes next to a pictogram showing the speed limit and types of road users (see Fig. 1). Four questions were asked: two concerning the road before the intersection and two concerning the road after the intersection:

1. What speed limit would you expect on the road before the intersection?
2. Which types of road users would you expect on the road before the intersection?
3. What speed limit would you expect on the road after the intersection?
4. Which types of road users would you expect on the road after the intersection?
Figure 1. Answer options used in the experimental task.

The questions about the speed limit allowed one and only one answer to be chosen, for the questions regarding road users multiple answers were allowed. Responding was self-paced. The total experiment took approximately 20 min.

Figure 2. Schematic illustration of the sequence of events in each individual trial.

The following road sections were used (see also Table 3 for the variants of layout mentioned below with the corresponding numbers):

- **Through roads:** 1) single carriageway with a broken centre line marking filled with green: 1 and 2) single carriageway with a broken centre line marking: 3
- **Distributor roads:** single carriageway with a broken centre line marking: 2, 4, 6, 8
- **Access roads:** 1) asphalt road without marking: 5 and 2) asphalt road with edge marking: 7

The following intersections were used:

- **TR-DR/ DR-TR transitions:** 1) rural interchange (typical: 9) and 2) signalized junctions (less typical: 10)
- **DR-AR/ AR-TR transitions:** 1) priority junction (typical: 11) and 2) roundabout (less typical: 12)
Table 3. Types of road layouts, intersections and transitions used in the study

<table>
<thead>
<tr>
<th>Transitions between a TR and a DR</th>
<th>Transitions between a DR and an AR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual road sections</strong></td>
<td></td>
</tr>
<tr>
<td>1. distinguishable</td>
<td>5. distinguishable</td>
</tr>
<tr>
<td>![Image 1]</td>
<td>![Image 5]</td>
</tr>
<tr>
<td>2. less distinguishable</td>
<td>6. less distinguishable</td>
</tr>
<tr>
<td>![Image 2]</td>
<td>![Image 6]</td>
</tr>
<tr>
<td>3. typical</td>
<td>7. less typical</td>
</tr>
<tr>
<td>![Image 3]</td>
<td>![Image 7]</td>
</tr>
<tr>
<td>4. less typical</td>
<td>8. typical</td>
</tr>
<tr>
<td>![Image 4]</td>
<td>![Image 8]</td>
</tr>
<tr>
<td><strong>Intersections</strong></td>
<td></td>
</tr>
<tr>
<td>9. typical</td>
<td>11. typical</td>
</tr>
<tr>
<td>![Image 9]</td>
<td>![Image 11]</td>
</tr>
<tr>
<td>10. less typical</td>
<td>12. typical</td>
</tr>
<tr>
<td>![Image 10]</td>
<td>![Image 12]</td>
</tr>
</tbody>
</table>

1) TR: single carriageway with a broken centre line marking filled with green; 2), 4), 6) & 8) DR: single carriageway with a broken centre line marking; 3) TR: single carriageway with a broken centre line marking; 5) AR: asphalt road without marking; 7) AR: asphalt road with marking; 9) rural interchange; 10) signalized junction; 11) priority junction; 12) roundabout

2.3. Design

A mixed factorial design was used. All participants watched clips which varied on three aspects (within-subject variables): (1) type of transition: transitions between a TR and a DR and transitions between a DR and an AR; (2) type of intersection: typical and less typical; (3) type of road layout: distinguishable and less distinguishable. The presentations order was varied across participants (between-subjects variable): 1) transition from a higher to a lower order road and 2) transitions from a lower to a higher order road. In total 16 clips were used and each participant watched 8 of them. Table 4 gives an overview of all experimental conditions.
2.4. Dependent variables

Two dependent variables were used in this study: (1) the correctness of the speed limit and (2) the correctness of access restriction. They were constructed as follows:

Correctness speed limit
Question 1 and 3 (see section 2.2) were used for this variable. Each transition was scored zero (0), one (1) or two (2) based on whether or not:
- A transition was correctly recognized as respectively a transition from a higher to a lower order road or from a lower to a higher order road (1 point).
- The speed limit of both roads (before and after a transition) was correctly indicated (2 points)
– Both the speed limit of one of the roads (before or after a transition) was incorrectly indicated and the transition was incorrectly recognized as a transition from a higher to a lower order road or from a lower to a higher order road (0 points).

**Correctness access restriction**

Question 2 and 4 (see section 2.2) were used for this variable. Each road user on both roads (before and after a transition) that was correctly indicated as to whether or not allowed on the road was scored one (1 point). For each transition a maximum of 4 points could be scored.

### 2.5. Analysis

**Main effects**

Since the dependent variables were non-normally distributed, non-parametric tests (Wilcoxon Signed Ranks tests) were used to test hypotheses 1 and 2. The main effects were analysed separately for expectations concerning speed limit and expectations regarding the presence or absence of different road users. To obtain main effects of intersection type (typical versus less typical) on the expectations concerning speed limit (respectively access restriction), the correctness speed limit scores (respectively correctness access restriction scores) of four clips containing typical intersections were summed up and compared with the sum score of the clips showing less typical intersections. In a similar way the main effects of road layout type and transition type were obtained. The maximum score obtained for speed limit was 8 (4 clips* 2 points), for access restriction 16 (4 clips*4 points).

**Interaction effects**

A chi-square test for categorical variables was carried out to study interaction effects. The interaction effects were analysed separately for expectations concerning speed limit and those concerning the presence or absence of different road users. To obtain interaction effects the average correctness speed limit scores (maximum 2 points) and average correctness access restriction scores (maximum 4 points) were used.

### 3. Results

First the data file of participants who watched clips showing transition from a lower to a higher order road (i.e. DR-TR and AR-DR transitions) were compared with the file of the ones who watched clips showing transitions from a higher to a lower order road (i.e. TR-DR and DR-AR transitions). There were no systematic differences between the files so the two sets were combined to one file. It has to be mentioned that a TR-DR transition from now on stands for both a TR-DR transition and DR-TR transitions and a DR-AR transition relates to both a DR-AR and an AR-DT transition.
3.1. Expectations of speed limit

Typical intersection types (such as a rural interchange and priority junctions) had a positive effect on the expectations concerning speed limit of the roads before and after a transition ($Z = -4.92, p < .01$). Furthermore, a transition was better recognized when the two road sections were distinguishable than when they were less distinguishable ($Z = -7.16, p < .01$). The results also show that it is easier to recognize a DR-AR transition than a TR-DR transition ($Z = -5.33, p < .01$). Interaction effects demonstrate that especially the recognisability of TR-DR transitions is affected by the layout of road sections and intersections (see Fig. 3). For TR-DR transitions a change in speed limit was more often noticed when the roads were characterized by a distinguishable layout ($\chi^2 (2) = 36.16; p < .01$) compared to a less distinguishable layout. It also appeared that the recognisability of a TR-DR transition increased significantly when it contained the typical intersection type (i.e. rural interchange) as opposed to a less typical intersection (i.e. priority junction). Figure 3 shows that DR-AR transitions were in general better recognized than TR-DR transitions. Only when TR-DR transitions were characterized by both distinguishable road layout and typical intersection were they better recognized than DR-AR transitions.

![Figure 3. Mean scores for correctly indicated speed limit.](image)

3.2. Expectations of road users

The results show that the typical intersections elicited more correct expectations concerning road users than the less typical intersections ($Z = -3.78, p < .01$). Additionally, the participants were more often correct about the access restriction of TR–DR transitions than the access restrictions of DR–AR transitions ($Z = -10.40, p < .01$). No significant effect of the road layout was found. TR-DR transitions were better recognised in terms of access restrictions than DR-AR transitions, especially when the former contained a rural interchange rather than a less typical intersection type. No significant interaction effects were found for expectations concerning the road users allowed on the road (see also Fig.4).
The first objective of the current study was to identify the optimal conditions under which drivers recognise transitions from one road category to another. Specifically the role of the road layout in combination with the type of intersection in enhancing recognisability of transitions between rural road categories has been investigated. Two types of transitions were addressed in this study: TR-DR transitions and DR-AR transitions characterized either by distinguishable or less distinguishable layout of the road sections and by typical or less typical type of intersection. Results reveal that the recognisability of transitions is affected by both the road layout and the intersection type.

First of all, when the layouts of the road sections are distinguishable, it is easier for drivers to notice that speed limit applying to the road before a transition differs form the speed limit of the road after a transition. Specifically, TR-DR transitions are best recognized in terms of speed limit when a TR is equipped with a green centre line marking. As for DR-AR transitions, the absence of road markings on ARs enlarges the distinction between AR and DR. These findings confirm the results of the earlier mentioned study (Stelling-Konczak et al., 2011) investigating the role of the layout of the road sections themselves in enhancing recognisability of transitions. Secondly, transitions are better recognised in terms of speed limit when they comprise typical intersections (e.g. a rural interchange in the case of TR-DR transitions) than when less typical intersections (e.g. a roundabout in the case of DR-AR transitions) make up a transition. Furthermore, the effect of the layout of the road section in combination with the intersection type was the most profound in the case of TR-DR transitions.
transitions. As for DR-AR transitions the differences were much smaller. However this type of transitions was already better recognized than TR-DR transitions.

Typical intersection types also help drivers to distinguish road categories in terms of the road users allowed on the road. Road layout, however, was here of no influence. Reverse to the speed limit effects, TR-DR transitions are better recognized in terms of access restrictions than the DR-AR transitions. Drivers are especially good at deciding which road users are allowed on the roads making up a TR-DR transition and especially bad at deciding which speed limit applies to roads making up a DR-TR transition. Furthermore, it appears that more mistakes concerning the estimations of speed limit are made in the case of TR-DR transitions and more mistakes regarding access restrictions are made in the case of DR-AR transitions. A likely explanation for the mistakes in speed limit concerns the similarity of the road layout of a TR and that of a DR as already mentioned in the introduction. The only characteristic on basis of which a distinction between TRs and DRs can be made is a difference in the type of the edge marking: a characteristic hardly noticed by road users. Given the identical separation of driving direction (both a TR and a DR has a centre line marking) it is more difficult to notice a change in a road category than when a separation of the driving direction is lacking on one of the roads (as in the case of DR-AR transitions). Consequently TRs and DRs are often wrongly seen as having the same speed limit. The phenomenon of change blindness can thus occur: since 80 km/h is the default speed limit on Dutch rural both DRs and TRs tend to be categorized as 80 km/h roads, irrespective of the edge line marking (Stelling-Konczak et al., 2011).

As far as the access restrictions are concerned, DRs and TRs are more alike than DRs and ARs. On both TRs and DRs slow traffic is not allowed. Agricultural vehicles are only allowed on a DR when alternative routes are lacking. With regard to ARs all traffic, including cyclists, moped riders and agricultural vehicles, is allowed. This implies that due to the similarity in access restrictions incorrect expectations concerning the speed limit of roads making up a TR-DR transition does not necessarily have to lead to wrong expectations regarding the road users allowed on these roads. This advantage does not apply to DR-AR transitions. Recognition of this type of transitions in terms of access restrictions seem to be more dependent on the correct expectations of the speed limit of the roads before and after a transition. The layout of the road sections and intersection type may positively affect the expectations of the speed limit of this type of transitions but one may subsequently not know which type of road user should be allowed on a road with this particular speed limit. To sum up, recognisability of a transition in terms of the expected speed limit does not mean the same as the recognisability in terms of the expected access restrictions, although both profit from the typical intersections types.
4.1. Limitations

As with every study, also this study had some limitations that have to be discussed. Firstly, this study used video clips to be watched while the real driving task involves much more than just viewing road environment in front of the car. We assumed that the animation clips represent to a certain degree real-life driving situations. Even within the limitation of this research method some transitions were better recognised than others which allows us to draw conclusions based on the relative differences. Secondly, expectations are also influenced by characteristics other than those varied in this study, such as type of pavement and road environment characteristics (e.g. Martens, 2007). Since the aim of this study was to investigate the role of the road and intersection layout we did not include these other characteristics in our experiment. Finally, the questions concerning both a road before and after a transition were asked after each clip. One could argue that the road category seen last has been better remembered and the expectations regarding a road after a transition are thus more often correct than the expectations concerning a road before a transition. Since the presentation order did not affect the results (see Results) we can however assume that the sequence of events in our experiment had no systematic effect on the results.

4.2. Implications and recommendations

Firstly, the present study emphasizes the importance of distinguishable road layout in enhancing recognisability of transitions between rural road categories. Although the green marking on a TR and an absence of edge marking on an AR enlarge distinction between these road categories and a DR, a few issues should be taken into account before recommending these characteristics for implementation. To start with, although the green marking has a positive effect on the recognisability of TR-DR transitions, it is not self-explanatory. Namely, the meaning of a colour does not necessarily have an intuitive link with functionalities such as a speed limit or particular road users to be expected on the road. In that case the green marking can only be helpful provided that the road users are well informed about its meaning (Stelling-Konczak et al., 2011). What is more, green marking on a TR does not fit the homogeneity principle. At speeds of above 70 km/h, Sustainable Safety promotes the use of some physical form of directional separation, especially on rural roads (Wegman et al., 2008). This in order to prevent head-on collisions which can have serious impact at high speeds. Applying the homogeneity principle uniformly would mean that both TRs and DRs should be equipped with physically separated driving directions as the speed limit of both TRs and DRs exceeds 70 km/h. In this case it is advisable to design each road category uniformly with its own unique separation of driving directions, ensuring that the DR and TR-variants are mutually exclusive. This will help drivers to distinguish the road categories. In case a physical separation of driving directions is not feasible, the green centre line marking constitutes a good alternative since it improves the distinctiveness of TR-DR transitions. As it cannot prevent driving on the opposite lane, it may be rather implemented as an interim solution.
towards a sustainable safe design. The same holds for the use of centre line marking on DRs and TRs in general, which is even more problematic as it provides no safety benefits and it undermines recognisability as well. Second, as stated above, from the point of view of a recognisable road layout, ARs should preferably not be provided with lineation. The results of this study also show that the transitions between a DR and an AR road are relatively well recognised both with and without the edge marking. From a road safety perspective a layout with edge marking is preferable even if it is somewhat less recognisable. The edge marking can namely improve safety during darkness or poor visibility by providing some cues about the road course.

When the layout of the road sections turns out to be insufficient to mark a transition between road categories, the typical intersection types can be used to enhance recognisability. In general, typical intersection types should rather be used than the less typical ones. A rural interchange is thus preferable to a signalized junction in case of TR-DR transitions. For DR-AR transitions however, the type of intersection is of less importance. These transitions are relatively well recognized and the use of a typical intersection: a priority junction instead of a roundabout does not help a lot more. Since roundabouts have demonstrated the potential to significantly reduce the most injurious (angle) type of crashes (Elvik, 2003; Lord et al., 2007), they are preferred as to priority junctions. Roundabouts are safer because they slow the operating speed of all vehicles, reduce the potential number of conflict points between the traffic movements passing and modify the most likely collision angles between vehicles (Elvik, 2003; Lord et al., 2007).

This study shows that a recognisable and predictable road layout can help drivers in forming correct expectations about transitions between rural road categories. At the same time the results of this study present a discrepancy between what drivers expect and recognise (i.e. green centre line marking on TRs, no edge marking on ARs) and what is safe for other reasons. It is thus important to keep defining essential characteristics of road categories while incorporating all sustainable safety principles into the ultimate road design, and not to content oneself to Essential Recognisability Characteristics. Only then a fully sustainable traffic system can be achieved.
REFERENCES


ECONOMIC COSTS OF ROAD TRAFFIC ACCIDENTS IN GERMANY
INTRODUCTION

Every year about 4,000 people die in traffic accidents in Germany. This means on average every one and a half hours a person loses his or her life in a traffic accident. Beside the loss of life serious physical and psychic traumata are another serious consequence of traffic accidents. Accidents have thus long-term impact on the life of the victims and their families alike. Hence, the fight against traffic accidents and their negative effects is a socio-political task of high priority. Beside the ethical and social motivation for the reduction of accident numbers and consequences, another aspect is that accidents tie up or even destroy personnel and material resources of the economy.

Transport policy requires economic analyses to support the decision making process and therefore needs monetary values which refer to and reflect the consequences of traffic accidents. Since the determination of these values is complex the resulting values are extrapolated for many years even though the validity of the values is strongly dependent on the economic and political circumstances. The last determination of the economic costs of road accidents was made in the middle of the 1990s. Basic conditions for the cost calculation have since changed dramatically.

In the light of these developments and the changes in the basic conditions it was necessary to reassess the costs of traffic accidents in Germany and to update the calculation methodology. This inquiry was carried out within a research project of the Federal Highway Research Institute with the objective to adapt the methodological approach as well as the basic input data to the current socio-economic conditions. Therefore the purpose of this article is to show the results of the empirical survey. However, because economic costs of road accidents cannot be observed directly the article wants to give an insight in the German methodology to assess these costs, which can only be estimated. [Baum/Kranz/Westerkamp 2010].

ACCIDENT COSTS AND IMPORTANT CHANGES IN BASIC CALCULATION PARAMETERS

Accident cost components

The assessment of the accident costs in Germany refers to different cost components which cover the costs resulting from the consequences of a road accident. In general the German accident cost calculation differentiates between two types of damages. First, personal damages are those damages which result from the fact that people are injured or killed in traffic accidents. Second, the damages to property result from the fact that objects are damaged or destroyed by accidents. Both damage types are subdivided in different cost components. Figure 1 shows the structure of the personal and property damages and the related cost components [Krupp/Hundhausen 1984, Baum/Höhrnscheid 1999, Baum et al. 2000, Baum/Kranz/Westerkamp 2010]
- Reproduction costs are the costs for the resources which are used to the restore the status before the accident has happened. Within the cost component of reproduction costs the cost calculation distinguishes the direct reproduction costs which result immediately from the restoration (e.g., vehicle repair, medical and social rehabilitation of injured persons), and the indirect reproduction costs which originate from the restoration of the legal status before accident occurrence (e.g. police costs, insurance administration costs, court costs).

- Costs of resource losses are costs of lost economic output. These costs accrue because productive resources are destroyed or their efficiency is temporarily or permanently affected. Therefore resource losses are the costs of lost potential production in the economy due to road accidents.

- While the costs of resource losses refer to the production possibilities at institutionalized markets, the lost added value outside of these markets (outer market production losses) is a complementary cost component. Value added outside of institutionalized markets results from household production and unpaid work as well as illegally undeclared work. In the last German accident cost inquiry these costs were analysed only for personal damages. However, vehicles are also used in the outer market production (e.g. shopping, material transports from DIY stores). Consequently this cost component was additionally included in the assessment of property damages.

- A decrease in people's productivity (victims and their relatives) can also originate from social and psychological consequences of an accident. This effect (e.g. due to psychological stress or the necessity to adjust personal plans for the future) is represented by the cost component of the humanitarian costs.

- Also, traffic jams caused by road accidents tie up productive resources because vehicles and people spend unproductive time in the traffic jam. Thus costs of lost travel time are also included in the accident cost assessment.
Although the cost components of property damages and personal damages are almost identical, the included effects and their economic assessments partly differ considerably. With the exception of the indirect reproduction costs and the costs of travel time losses, every single cost component had to be determined separately for personal damages and the damages to property.

**Update and amendment of the German accident cost calculation**

An accident cost assessment which is up to date with the current socio-economic conditions and also methodological comparable in the European context is essential for an efficient road safety work in Germany and Europe. Nevertheless, since the last analytic-empiric inquiry in the middle of the 1990s knowledge about accident cost assessment has changed. This applies both to the inquiry and the assessment methodology as well as the actual conditions in the restoration and the production processes in the economy.

Furthermore basic conditions for the accident cost calculation in Germany have changed significantly. In particular the so called agenda 2010 included reforms of the social system (e.g. pensions, health insurance) which partially have a considerable impact on the economic costs of road accidents. For example the range of available health services was reduced by the law for the modernisation of the compulsory health insurance. Also the default retirement age was raised from 65 to 67 years within the pension reform.

Due to these developments there was a serious need to update the accident cost calculation in Germany. This update included the following aspects [Baum/Kranz/Westerkamp, 2010]:

- For all cost components new calculation models had to be developed.

- Due to the changed basic conditions an empirical survey about the input data for the cost assessment was necessary. The year 2005 was chosen as reference year because officially extrapolated accident cost values of the Federal Highway Research Institute existed up to the year 2004. Additionally, data about accidents and their consequences (which may last several years, e.g. medical follow-up treatments) are not available directly after the accident has happened. The probability to consider all relevant costs of an accident in a survey increases with the time distance to the chosen reference year.

- As already mentioned earlier investigations of accident costs did not include travel time delays due to accidents. To close this gap one objective of the accident cost assessment was to estimate the time delays due to road accidents with a calculation model and to assess the value of the travel time losses with suitable assessment methods.

- The calculation of the accident costs of personal damages is differentiated according to injury severity. The official German road accident statistics differentiates three
categories: Fatality, seriously injured person, slightly injured person. Besides these categories the costs of an additional category “very seriously injured” have been analysed. The reason for this analysis was the suspicion that the steady decline of the number of fatalities in Germany goes along with a growing proportion of persons with more severe injuries within the category of the serious injuries.\(^1\) Therefore the “very seriously injured” is a subcategory of “seriously injured” containing personal damages with very severe injuries which increase the likelihood to suffer permanently from the consequences of an accident. To estimate the proportion of the “very seriously injured” within the category of the “seriously injured”, the official road accident statistic is not sufficient [Höhnscheid 2005 and Lefering 2007]. Consequently an estimation of the proportion of the “very seriously injured” was based on a complementary inquiry based on additional, available data sources. For this inquiry it was necessary to combine the cost data of the survey with the additional data source. This connection between the data sources required to set criteria to identify cases of the empirical survey of the accident cost inquiry which can be assigned as very seriously injured.

- Accident statistics have to cope with the problem of underreporting, i.e. accidents are partially not recorded in the official statistics. A study of ICF Consulting from 2003 calculated an average rate of underreporting of approximately 30 percent for EU accident statistics [ICF 2003]. Due to the legal situation in Germany no underreporting of personal damages was to be expected. However, it is well known that the problem of underreporting of property damages is the same in Germany as in the international context. For the accident cost calculation it is important to know which accident categories are affected by underreporting and how this affects the accident costs. Even though underreporting was not expected for the personal damages it had also to be evaluated if there was any evidence about an underreporting of personal damages and if there was, how this would affect the accident costs.

EMPIRICAL ACCIDENT COST SURVEY

Beside available general statistical data the accident cost inquiry required raising additional input data in an empirical survey. This survey was made in cooperation with institutions which have to pay for the consequences of road accidents such as insurance companies. For this purpose a sample of accidents was raised by the evaluations of randomly chosen cases (accident records) at the insurance companies (car insurances, statutory health insurances and the Statutory accident insurances). Several associations and insurers have moreover undertaken special evaluations for this investigation including, for example, the portion of lawyer’s costs in case of a legal dispute. The data from the sample was used to calculate the

\(^1\) In Germany severe injured persons are defined as persons who need to stay in hospital for about 24 hours due to their accident related injuries (and do not die within 30 days after the accident, so they were counted as Fatality).
reproduction costs and the costs of resource losses (especially information about the medical costs or the duration of medical treatment).

The data availability varied between the institutions. Therefore not all data could be raised in every supporting institution so for some data it was necessary to raise a larger sample at those institutions which are in charge of rarely available data (this especially concerns information about disability and fitness for work which are only known by statutory accident insurances). The accident records which contained no cost data were not considered in the survey. Table 1 gives an overview about the structure of the raised data set.

Table 1: Required data-set for the empirical survey of accident costs [Baum/Kranz/Westerkamp 2010]

<table>
<thead>
<tr>
<th>Data about personal damages</th>
<th>Data about property damages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institution</td>
<td></td>
</tr>
<tr>
<td>Injury severity degree (fatal, seriously injured, slightly injured)</td>
<td>Accident category</td>
</tr>
<tr>
<td>Date when accident happened</td>
<td></td>
</tr>
<tr>
<td>Accident location (motorways, rural areas, urban areas)</td>
<td></td>
</tr>
<tr>
<td>Gender of injured person</td>
<td>Number and type of damaged vehicles</td>
</tr>
<tr>
<td>Year of birth of the injured persons</td>
<td>Accident reported to the police</td>
</tr>
<tr>
<td>Duration of temporary disability</td>
<td>Costs for repair or replacement</td>
</tr>
<tr>
<td>Duration of outpatient treatment</td>
<td>Reduction in vehicle value</td>
</tr>
<tr>
<td>Duration of hospital treatment</td>
<td>Vehicle residual value</td>
</tr>
<tr>
<td>Costs of ambulance services</td>
<td>Other property damages</td>
</tr>
<tr>
<td>Costs of outpatient treatment</td>
<td>Other costs caused by the property damage (e.g. costs for towing or registration of replacement vehicle)</td>
</tr>
<tr>
<td>Costs of ambulatory rehabilitation</td>
<td>Time in which the vehicle could not be used</td>
</tr>
<tr>
<td>Costs of medical aid/equipment</td>
<td></td>
</tr>
<tr>
<td>Costs of vocational rehabilitation</td>
<td></td>
</tr>
<tr>
<td>Duration of inpatient rehabilitation</td>
<td></td>
</tr>
<tr>
<td>Costs of inpatient rehabilitation</td>
<td></td>
</tr>
<tr>
<td>Duration of hospital treatment</td>
<td></td>
</tr>
<tr>
<td>Reduction of fitness for work (percent)</td>
<td></td>
</tr>
<tr>
<td>Duration of reduction of fitness for work</td>
<td></td>
</tr>
<tr>
<td>Costs for in- and outpatient care</td>
<td></td>
</tr>
<tr>
<td>Need for care (defined by law)</td>
<td></td>
</tr>
<tr>
<td>Funeral expenses</td>
<td></td>
</tr>
</tbody>
</table>

The survey for personal damages involved the analysis of the costs of about 9,000 injuries (damages). Partly the survey was made by manual evaluation of accident records at the insurance companies, partly by IT-based evaluation of the insurances. An IT-based special evaluation of the German Statutory Accident Insurance (DGUV) covered more than 184,000 injuries damages and focused on cost data and fitness for work.

The sample for property damages exclusively involved data from car insurance companies gained by manual evaluation of accident records. The calculation is based on an assumed normal distribution of the sample and a confidence level of 95 percent.
Compared to the actual sample sizes in the survey (listed in Table 2 and Table 3), the sample error for personal damages for all injury severity degrees is less than two percent. A further reduction of the sample error would have required a disproportionate rise of the sample size.

**Table 2: Actual sample size of the survey and the special evaluation [Baum/Kranz/Westerkamp 2010]**

<table>
<thead>
<tr>
<th>Injury Severity</th>
<th>Sample size of the survey (car insurance, health insurance)</th>
<th>Sample size of the special evaluation (statutory accident insurance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatal injuries</td>
<td>144</td>
<td>681</td>
</tr>
<tr>
<td>Serious injuries</td>
<td>2.272</td>
<td>20.686</td>
</tr>
<tr>
<td>Slight injuries</td>
<td>7.053</td>
<td>163.174</td>
</tr>
</tbody>
</table>

For the damages to property the sample error is above 5 percent for accidents with fatalities, above 7 percent for accidents with slight injuries and 11 percent for accidents with serious injuries. Another reduction of the sample error in the survey by increasing the sample size was not possible because the car insurances were not able to select accident records by damage type. Nevertheless, the actual sample size of the survey enables a valid analysis of personal damages. Moreover, the calculation of property damages is based on data about the cost estimation in the police report. The sample of the survey enables a foundation of the property damage calculation by comparing the result to the estimated values.

**Table 3: Actual sample size of the survey of property damages [Baum et al. 2010].**

<table>
<thead>
<tr>
<th>Category</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accidents with fatal injuries</td>
<td>103</td>
</tr>
<tr>
<td>Accidents with serious injuries</td>
<td>52</td>
</tr>
<tr>
<td>Accidents with slight injuries</td>
<td>157</td>
</tr>
<tr>
<td>Severe accidents with only property damage</td>
<td>164</td>
</tr>
<tr>
<td>Other accidents with only property damage</td>
<td>310</td>
</tr>
</tbody>
</table>

Beside data about the medical treatment, the special evaluation of the German SAI also provided information about decreased fitness for work (the so called MdE-Scale measures this reduction in percent.). All injuries with a permanently decreased fitness for work recorded for the years 2004 to 2007 by the statutory accident insurance were evaluated. All in all this includes more than 11,890 injuries. This sample size ensures a valid inquiry of reduced fitness for work [Baum/Kranz/Westerkamp, 2010].

**Inquiry of the personal damage costs**

The costs of the personal damages result from the injuries or the death of accident victims. As shown in figure 2, the reproduction costs originate from the medical treatment, the costs of resource losses and the outer market costs which result from the time accident victims are not available in the production process. However, the indirect reproduction costs arise from the restoration of the legal status before the accident has happened. Figure 2 gives an overview
about the components and input data of the personal damage cost calculation. For the calculation of the economic costs of personal damages it was necessary to elaborate calculation models for each cost component.

**Figure 2: Economical costs of personal damages [Kranz 2010]**

**Direct reproduction costs.** The calculation of the direct reproduction costs is based completely on the accident cost data raised by the insurances.

**Indirect reproduction costs.** The calculation of the indirect reproduction costs is partially based on data of official statistics, as for example the judiciary statistics of the German Federal Statistical Office. However, gaining additional data was necessary (study results from other fields of research (e.g. BMFSFJ 2005 for costs of job replacements and specific information from branch associations and companies).

**Costs of resource losses.** The calculation of costs due to resource losses assesses lost potential production because of injuries or the death of the accident victims. The amount of time the victims are not available in the production process was calculated on the basis of the survey data for the duration of medical treatment and the decrease of fitness for work as well as the statistical knowledge about the age of the accident victims. The monetary assessment of the time losses was carried out following the Cobb Douglas Production function of the German Central Bank [Deutsche Bundesbank 2003]. The Production function enables an estimate of a relative change of the potential GDP that can be determined by the change of the productive factors (labor and capital) The determined relative change was put in relation to the potential GDP estimate of the German Council of economic experts [Council of economic experts 2007]:

\[
\Delta \ln Y = \Delta \ln A + \alpha \cdot \Delta \ln L + (1 + \alpha) \cdot \Delta \ln K
\]

With:
\[ \Delta \ln Y \] Change in potential Gross Domestic Product (GDP)

\[ \Delta \ln A \] Change in total factor productivity (technical progress)

\[ \Delta \ln L \] Change rate of the productive factor labor

\[ \Delta \ln K \] Change rate of the productive factor capital

\[ \alpha \] Factor elasticity

*Outer market production losses.* Lost added value in the informal economy (illicit work) was calculated as a simple percentage of the costs of resource losses. To determine this percentage results of a study on illicit work in Germany were used. According to the study the added value of illicit work is about 15.4 percent of the total GDP [Ernste/Schneider 2006].

The determination of lost time for household production and unpaid work was accomplished according to the assessment of costs of resource losses. As a monetary value the average added value of the activities outside of institutionalised markets has been used. This value has been assessed in the time budget elevation of the Federal Statistical Office of Germany [Schäfer 2004, Baum / Kranz / Westerkamp 2010].

*Humanitarian costs.* The inquiry of the humanitarian costs is based on an evaluation of court decisions about compensation payments. The assessment is based on 705 court decisions. The humanitarian costs enclose exclusively consequences related to resource consumption. Hence, the calculation of an intangible risk value (as done in the international context) was not included in the German accident cost calculation.²

**Inquiry of the property damage costs**

The direct reproduction costs of property damages enclose the costs which are related to the repair or the replacement of the damaged objects. More in detail the damages to property enclose the vehicle damages but also damages to other objects such as fences or trees. In contrast to the personal damages the indirect reproduction costs contain a lower number of cost components (no job replacement costs and funeral expenses). Figure 3 gives an overview about the cost components and necessary input data

---

² An evaluation of the risk value on basis of a adjusted European average value was carried out for comparative purposes [Bickel et al. 2005]. Including the risk value in the German accident cost calculation would result in an increase of the costs per fatality of about 1.32 million Euros (130 percent) [Baum/Kranz/Westerkamp 2010].
Direct reproduction costs. The calculation of the direct reproduction costs due to property damages is based on reported estimates of the police [Baum et al. 2000]. Because of the deviation of these estimates from the actual costs the estimates were corrected with a correction factor (Hautzinger et al. determined different factors in an empirical inquiry). All property damage costs which have not originated from vehicle damages were calculated exclusively at basis of the data of the empirical survey.

Indirect reproductive costs: The corresponding indirect reproduction costs components of the personal damages and property damages property were calculated in joint models for each component. The total indirect reproduction costs were allocated to personal and property damages.

Costs of resource losses: The costs of resource losses were calculated analogously to the methodology used for the personal damages. Average downtimes (time the damage vehicle could not be used) were determined and assessed monetary.

Outer market resource losses: As for personal damage lost value added in informal markets were assessed for property damages. For illicit work it was calculated according to the methodology for personal damages with a percentage of the resource losses. Nevertheless, for the lost added value from housework production and unsalaried work the value of the availability to use a vehicle for these activities has been determined. The costs of vehicle downtimes have been calculated using the official statistics about registered vehicles combined with official lists for compensation payments for vehicles which are not available due to damages (both the lists and the statistics contain information about different vehicles).
Inquiry of the travel time losses caused by accidents

An evaluation of the costs of time delays caused by road accidents on federal motorways was carried out on the basis of traffic data of the year 2005 [SVZ 2005]. Using a deterministic queuing model, assumptions about the capacity restriction caused by accidents [Listl et al. 2007] and accident rates depending on the traffic volume [Pöppel-Decker/Schepers/Kossmann 2003] average travel time delays were estimated for accidents. The calculated time losses were valued monetarily. Figure 4 gives a schematic overview about the calculation of travel time delays caused by traffic accidents.

![Figure 4: Calculation of the travel time delays caused by accidents](image)

The monetary assessment of the travel time delays was executed analogously to the methodology of the costs of resource losses and the outer market resource losses. Time delays due to accidents were interpreted as a decrease of the potential production. The costs of the time delays in 2005 amount to 264.66 million euros. These costs are only caused by time delays caused by accidents costs and do not include other capacity-limiting factors like weather or construction sites.

RESULTS OF THE COST INQUIRY

Results of the Basic Cost Calculation

All together the total economic costs of road accidents of the year 2005 amount to 31.477 billion Euros. Personal damages caused costs of 15.226 billion Euros (48.37 percent) and 16.251 billion Euros (51.63 percent) are originated by property damages. Table 4 shows the proportion of the different injury severity degrees on the total costs. Table 5 shows the accident costs per injury. The total costs of the different accident categories of the property damages are listed in Table 6. Table 7 provides an overview about the costs per accident.
### Table 4: Personal damages in 2005 in million euros

<table>
<thead>
<tr>
<th></th>
<th>Fatal injuries</th>
<th>Serious injuries</th>
<th>Slight injuries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct reproduction costs</td>
<td>25,255</td>
<td>1,096.226</td>
<td>100,691</td>
</tr>
<tr>
<td>Indirect reproduction costs</td>
<td>125,242</td>
<td>848.321</td>
<td>307,246</td>
</tr>
<tr>
<td>Costs of resource losses</td>
<td>2,795.336</td>
<td>3,100.677</td>
<td>298,255</td>
</tr>
<tr>
<td>Costs of outer market resource losses</td>
<td>2,342.911</td>
<td>2,126.584</td>
<td>132,440</td>
</tr>
<tr>
<td>Humanitarian Costs</td>
<td>169,099</td>
<td>944,858</td>
<td>695,956</td>
</tr>
<tr>
<td>Costs of time delays on federal motorways</td>
<td>3,026</td>
<td>29,077</td>
<td>84,312</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>5,460.871</strong></td>
<td><strong>8,145.742</strong></td>
<td><strong>1,618,900</strong></td>
</tr>
</tbody>
</table>

### Table 5: Costs per injury in 2005 in euro

<table>
<thead>
<tr>
<th></th>
<th>Fatal injuries</th>
<th>Serious injuries</th>
<th>Slight injuries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct reproduction costs</td>
<td>4,710.92</td>
<td>14,245.58</td>
<td>282.45</td>
</tr>
<tr>
<td>Indirect reproduction costs</td>
<td>23,361.65</td>
<td>11,024.02</td>
<td>861.86</td>
</tr>
<tr>
<td>Costs of resource losses</td>
<td>521,420.61</td>
<td>40,293.65</td>
<td>836.64</td>
</tr>
<tr>
<td>Costs of outer market resource losses</td>
<td>437,028.73</td>
<td>27,635.19</td>
<td>371.51</td>
</tr>
<tr>
<td>Humanitarian Costs</td>
<td>31,542.59</td>
<td>12,278.53</td>
<td>1,952.24</td>
</tr>
<tr>
<td>Costs of time delays per injury on federal motorways</td>
<td>(4,572.22)</td>
<td>(4,961.18)</td>
<td>(3,180.88)</td>
</tr>
<tr>
<td><strong>Total (without costs of time delays)</strong></td>
<td><strong>1,018,064.51</strong></td>
<td><strong>105,476.98</strong></td>
<td><strong>4,304.70</strong></td>
</tr>
</tbody>
</table>

### Table 6: Damage to property costs per accident in 2005 in euro

<table>
<thead>
<tr>
<th></th>
<th>Accidents with fatal injuries</th>
<th>Accidents with serious injuries</th>
<th>Accidents with slight injuries</th>
<th>Serious accidents with only property damage</th>
<th>Other accidents with only property damage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct reproduction costs</td>
<td>22,501.30</td>
<td>10,765.13</td>
<td>6,898.18</td>
<td>10,168.26</td>
<td>2,854.68</td>
</tr>
<tr>
<td>Indirect reproduction costs</td>
<td>13,878.18</td>
<td>6,500.47</td>
<td>4,509.06</td>
<td>6,306.52</td>
<td>1,772.42</td>
</tr>
<tr>
<td>Costs of resource losses</td>
<td>1,482.70</td>
<td>771.85</td>
<td>605.15</td>
<td>915.10</td>
<td>372.13</td>
</tr>
<tr>
<td>Costs of outer market resource losses</td>
<td>481.54</td>
<td>348.83</td>
<td>314.02</td>
<td>493.95</td>
<td>278.29</td>
</tr>
<tr>
<td>Costs of time delays</td>
<td>3,139.58</td>
<td>4,054.37</td>
<td>3,225.70</td>
<td>6,928.57</td>
<td>-</td>
</tr>
<tr>
<td><strong>Total costs of property damages per accident</strong></td>
<td><strong>38,343.72</strong></td>
<td><strong>18,386.27</strong></td>
<td><strong>12,326.42</strong></td>
<td><strong>17,883.82</strong></td>
<td><strong>5,277.53</strong></td>
</tr>
</tbody>
</table>
Table 7: Damage to property costs in 2005 own calculation in million euro

<table>
<thead>
<tr>
<th></th>
<th>Accidents with fatal injuries</th>
<th>Accidents with serious injuries</th>
<th>Accidents with slight injuries</th>
<th>Serious accidents with only property damage</th>
<th>Other accidents with only property damage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct reproduction costs</td>
<td>112.15</td>
<td>717.25</td>
<td>1,828.07</td>
<td>1,017.57</td>
<td>5,187.81</td>
</tr>
<tr>
<td>Indirect reproduction costs</td>
<td>69.17</td>
<td>433.11</td>
<td>1,194.94</td>
<td>631.11</td>
<td>3,221.02</td>
</tr>
<tr>
<td>Costs of resource losses</td>
<td>7.39</td>
<td>51.43</td>
<td>160.37</td>
<td>91.58</td>
<td>676.28</td>
</tr>
<tr>
<td>Costs of outer market resource losses</td>
<td>2.40</td>
<td>23.24</td>
<td>83.22</td>
<td>49.43</td>
<td>505.74</td>
</tr>
<tr>
<td>Costs of time delays</td>
<td>1.86</td>
<td>17.82</td>
<td>51.68</td>
<td>116.89</td>
<td>-</td>
</tr>
<tr>
<td>Total costs of property damages</td>
<td>192.96</td>
<td>1,242.84</td>
<td>3,318.28</td>
<td>1,906.58</td>
<td>9,590.85</td>
</tr>
</tbody>
</table>

Results of the Complementary Investigations

The so far presented results exclusively refer to the cases of damages reported in official road accident statistics (accidents and injuries). Two areas which are out of the scope of the official statistics were examined complementary:

**Very seriously injured** – For the first time an inquiry of the economic costs of very seriously injured persons showed that these injuries with long term consequences are related to 37-times higher costs (approx. 395,000 Euro per injury) than average injuries due to accidents. This result illustrates the urgent need for action in this field of road safety work.

**Underreporting** – The effects of a consideration of unreported accidents on the total accident costs in Germany was analyzed by comparing if accidents reported to insurances have also been reported to the police. The result of the analysis was that unreported property damages would result in additionally 7.04 billion Euros. Primarily this increase goes back to minor damages (other accidents with property damage only). Unreported slight injuries were related to a cost increase of about 0.41 billion Euro, while there were no observed unreported fatalities or serious injuries.

**DISCUSSION**

Road Accidents are a negative consequence of a modern society. An increasing demand for transport and mobility results in the deaths and injuries of humans and damages to property. The only way to avoid road accidents completely would be, to do without all forms of transportation. Easy to see, that this would not only be a significant decrease of individual freedom but the end of a society which is based on the division of labor. Road safety therefore always is a question of balancing the societal need to find a suitable degree of road safety.

However, although road safety can be increased by private investments, transport policy is responsible for many decisions concerning road safety. Therefore a socio-economic assessment is necessary which is based on the fact, that transport does not only increase productivity of resources but also consumes them.
The calculated costs of road accidents in Germany are essential for the socio-economic assessment of road safety measures in Germany and the EU. The values represent the impact of different consequences of road accidents to the economy. The values replace further calculated cost values and therefore will be widely used in public decision making processes of the automotive industry and road operators. For instance the values are used in the Safety analysis of road networks in Germany and the German Guideline for cost benefit studies of roads.

Although the calculated accident costs consider current socio-economic conditions, the aggregation of the values sets tight limits to a differentiated cost analysis in specific contexts (e.g. cost of motorcycling accidents) because the calculated accident costs do just describe the costs of average damages but do not reflect the specialities of specific accidents. Further efforts should be made to cost inquires of specific important accidents like accidents with motorcyclist, vulnerable road users or heavy goods vehicles.

A more disaggregated cost analysis in decision making also needs to give more detailed evidence about the severity of injuries. The cost inquiry showed that very severe injuries are related to 37-times higher costs than other severe injuries (with treatment in hospital). However, the calculated values just give an insight into the short term costs of these injuries but due to their severity long term consequences should also be taken into account. Further research is necessary to reveal the long term consequences and costs of very severe injuries.

The calculated economic costs of road accidents in Germany include ex-post values. These values do not reflect the ex ante value of a risk reduction in road safety which is usually estimated in stated-preference-surveys. Ex ante costs may have a major impact on the costs of road accidents, for example a current Norwegian surveys shows, that the relative importance of the risk value amounts from 49.36% for serious injuries up to 86.45 % for fatalities [Veisten / Flügel / Elvik, 2010]. The assessment of German road safety measures does not include the ex ante costs, because of a compatibility to the monetary value of other effects of transport. Thus the German values are not directly comparable to the economic costs of road accidents in other countries. The calculation reflects the specific socio economic conditions in Germany and do not include an ex ante valuation of a risk reduction.

**CONCLUSION**

Costs of road accidents in Germany (including costs of time delays due to accidents on federal motorways) amount to 31.48 billion Euro which is equivalent to 1.4 percent of Germanys GDP of the year 2005.

Personal damages caused costs of 15.226 billion Euros (48.37 percent) and 16.251 billion Euros (51.63 percent) are originated by property damages.
REFERENCES


GEOGRAPHICAL LOCATION OF DEPOPULATION AREAS IN THE CZECH REPUBLIC AND ITS DEPENDENCE ON TRANSPORT INFRASTRUCTURE

INTRODUCTION

One of the burning problems of the current European society is the uneven growth of regions connected with the concentration of development to central areas and the decline of remote, marginal areas. The economy of declining marginal regions is heavily supported by subsidies from development funds, but these costly measures do not often ensure the expected benefits. Marginalization of these regions is a complex social phenomenon related to changes in lifestyle of the society, as well as to demographic changes, which are only handled with difficulties. Since this phenomenon is in a large part influenced by geographical characteristics, particularly the parameters of distance and transport accessibility of these regions from centres, a question is arising how transport and transport infrastructure influence such a phenomenon and how they may change it. Factors of economic and time demanding commuting to work, accessibility, frequency and quality of public transport, density and quality of roads are all factors greatly affecting everyday reality of life in rural areas. How are then these factors affecting the attractiveness and non-attractiveness of a given village or a region? This article aims to answer this question.

MARGINALIZATION AND DEPOPULATION IN RURAL AREAS

Marginalization is a phenomenon related to economic decline of a certain area with simultaneous decline in the importance of this area for economic and social activities, which lead to its loss of attractiveness as a place of residence or a recreation area (see Drápela, 2010). Marginalization is manifested in many spheres, and Leimgruber (2004) considers the economic, socio-cultural, political, and environmental marginality as crucial. In real life, the developing process of marginalization is manifested in the loss of population in exposed areas, when this loss is caused by both the negative balance of migration, and the natural decrease of population due to population ageing. The reasons for marginalization of certain areas are explained differently, e.g. Galante and Sala (1987) maintain, that the marginalization is caused by certain disadvantages (as location,
orographic situation etc.) of these regions, taking into account different economic and social indicators, which subsequently make them uncompetitive. Andreoli (1992) and Schmidt (1998) then search for the reasons of marginalization in insufficient integration into structures, processes, and systems which are dominant in a given place and time. The position outside the dominating structures and systems is considered by the theory of mainstream (Giddens, 1984), which states that the marginalization affects those units which are different from the majority in crucial economic and social parameters. Tykkyläinen (1998) then understands the marginalized area as a borderline area of socio-economic activity, i.e. at the edge of the socio-economic system. Mehretu et al. (2002) claim that marginalization may be caused by more different causes, while they distinguish:

- contingent marginality, which is a result of free market with equal competition, when the negative results of the competition come from the competition inequality.
- systemic marginality, effective mainly in totalitarian systems where the hegemonic powers of political and economic apparatus bring about discrepancies in the distribution of social, political, and economic benefits.
- collateral marginality, originating as an unexpected side effect of the process. It is a type of a neighbouring effect, when a member of the majority may be unintentionally marginalized for its closeness to a marginal minority.
- leveraged marginality, which is a result of an intentional process, when the pressure applied by economic players, requiring the highest profits and lowest costs, leads to marginalization.

Whatever reasons for marginalization in a given area, the typical result of this process is the depopulation of the given area (Drápela, 2010). In addition, the gradual loss of population subsequently intensifies marginalization of such region and generates other problems caused by the ever decreasing population density. This snowball effect may cause a long-term problematic situation, when such depopulated areas are only attractive concerning inexpensive residency, which attracts particularly underprivileged inhabitants, who in turn have difficulties in finding jobs due to their lower qualification. Therefore, marginalized areas may become regions with low population density with predominantly underprivileged population and with poor offer of jobs. To ensure this vision may not come true, it is necessary to reveal the marginalization of a region sufficiently in advance and prevent the depopulation of affected disadvantaged areas.

**METHODOLOGY**

Before introducing the methods used in this study, it is essential to specify more closely the research goals. Although in a dissertation, which was a model for this article, the number of objectives was higher, this article focuses on:

1. delimiting depopulated areas in the Czech Republic in post-communist era,
2. assessing the significance of the impact of transport infrastructure and commuting times on the current distribution of depopulated areas and,
3. assessing the significance of transport for the development of marginal areas in the long-term perspective.

The first point includes an indicator of population movement (or the total growth or decrease of population within an administration unit). The unit selected for the analysis was a municipality, of which there are 6251 in the Czech Republic. The level of municipalities was selected since it is the most detailed reasonable level on which it is still possible to evaluate quantitatively expressed population growth or decrease. Data would still be available for the level of municipal parts; however, some municipal parts have fewer than 10 inhabitants and thus each small change would have a huge effect in comparison with other values. Therefore, one step higher level, the municipality, was chosen.

The values of the indicator were compared for decades between individual censuses, taking place in the monitored period in 1991 and 2001; the most up-to-date census took place in March 2011. However, the final results from the latest census will only be presented in the following years. The data from the census of people, flats and houses were used for 1991 and 2001, whereas the continuous population data was used for 2010. The values of the indicator “population movement” were assessed in such a way that a decrease of population in these decades by more than 15 % was considered as “serious marginalization”, a decrease by 5 – 15 % as “important marginalization”, and a decrease below 5 % as “mild marginalization”. The last mentioned category was not paid so much attention to, since due to generation changes in municipalities, temporary changes of permanent residency of inhabitants, and temporary declines caused by migration, smaller municipalities may easily fall into this category, while being virtually non-marginalized.

The second objective, the assessment of significance of the impact of transport infrastructure and commuting times on the current distribution of depopulated areas, became the initial point for a correlation analysis. This analysis was supplemented by case studies results executed in selected depopulated areas in an even more detailed scale and supported by a socio-geographical survey among the population. The correlation analysis was executed based on data from the census in 2001 and the population movement indicator values were compared with values of the 17 most relevant indicators concerning education, population size of municipalities, population age structure, employment and unemployment, economic activity, commuting and its time terms. The complete list of indicators is specified in the corresponding section.

The correlation analysis results were then compared with qualitative data collected with questionnaire surveys in selected model areas, and with historical and geographical research, which are focused on finding impulses having significantly influenced the development of towns and villages in these regions in the past, regardless whether positively or negatively. Taking into account the size of the present article, it is impossible to present the complete methodology of the qualitative research. Nevertheless, it will be briefly outlined. The case
studies were performed in the model areas of the Nové Město region, for its location representing mountain and foothill regions, the Litenčice region, for its location representing richer agricultural regions, and the Pelhřimov region, for its special historical development when virtually the whole region was the property of one owner, the order of Premonstratensians, representing a region with a long-term planned population structure. The questionnaire survey was executed with regard to objective 3 in the above mentioned model areas with a sample of 100 inhabitants. The questionnaire was focused on both the subjective assessment of transport infrastructure in a given locality the objective being to understand the inhabitants’ perception and preferences, and the actual transport behaviour of the inhabitants, with a structure that was very similar to a typical travel survey. In the first part of the survey people answered questions concerning their subjective assessment of quality and density of transport infrastructure in place of their residence, localities which they view as potential threat to their safety in traffic, their preferences and reasons when choosing a transport mode, satisfaction with public transport services, their quality, routes, frequency and prices; the final open question concerned their recommendations for further development of transport and transport networks in the locality. The main outcome of this part was a SWOT analysis (strategic planning method used to evaluate the involved Strengths, Weaknesses, Opportunities, and Threats) of transport situation in the area. The other part of the questionnaire was used as a improvised travel survey when the respondents answered the questions concerning their real spatial mobility: why, where, how often, by which transport means, how long, how far, and which way they travel. Subsequently, based on these data, it was possible for each region to compile a set of schemes of a given municipality inhabitants’ spatial mobility, which supplements the SWOT analysis based on subjective impressions with an objective perspective.

Only marginally were used the results of the historical and geographical research, which was otherwise a substantial part of the dissertation, since it rather focused on different factors, particularly physio-geographical, ownership, historical, socio-cultural, economic, environmental, or a factor of externalities impact. Within this research with the use of historical sources, the development of municipalities in model areas from the date of the origin (e.g. already from the 11th century) until today was analysed: number of inhabitants and houses, records of economic activities, rights and privileges, disasters, and intentions of owners-feudalist, etc. up to civic amenities, political decisions having direct impact on given localities, and public transport. These data was then subjected to a critical assessment of their impact on development or decline of these municipalities, and the impact of individual factors is illustrated in examples. Regarding the factor of transport, the last 200 years, related to the industrialization and the development of road and railway network, are taken into account.
DEPOPULATION AREAS DEFINITION

As mentioned in the previous section, the definition of a depopulation area, i.e. an area with negative values of inhabitant mobility indicators, is completed with threshold values, derived from empirical experience, dividing the set of communities with a decrease of population into three groups: slightly affected by depopulation (decrease of inhabitants by up to 5 %), where this phenomenon is often only of short-term dimension, significantly affected by depopulation (decrease of inhabitants by 5 – 15 %) and considerably affected by depopulation (decrease of inhabitants over 15 %). The attention should thus be aimed at the two latter categories and at their relation to the spatial distribution of main roads.

The situation between 1991 and 2001 is depicted in Figure 1, the period between 2001 and 2010 in Figure 2. The black colour represents the main urban agglomerations, where the most important are the Prague agglomeration with about 1.5 million of inhabitants, and the Brno and Ostrava agglomerations, each with 0.5 million of inhabitants. The various shades of grey indicate the population rate.

The Figures show that the spatial distribution of depopulation areas in the Czech Republic depends especially on the distances of these regions from centres. The depopulation areas create something like rings around the main centres, interrupted only at locations where the big centres are not far away from one another. Within this phenomenon no considerable

Figure 1: Depopulation areas in the Czech Republic between 1991 and 2001 and their relation to spatial distribution of main roads
influence of the proximity to main roads (e.g. especially motorways) is visible as a factor significantly preventing depopulation. If a road runs in the proximity of a depopulation area, but the communities in the direct proximity of the road do not suffer from a decrease of inhabitants, it is rather caused by a different land relief, where the road runs through lowlands, whereas the depopulation area is situated in an highland or mountain terrain. In a situation of comparable orographic conditions in areas directly around the road as well as the ones farther from it, its positive influence on depopulation is usually not applied. On the contrary, outside these depopulation areas, in areas closer to centres, there exist isolated depopulation units or their clusters that differ from their prosperous surroundings by lower residence attractiveness. This is often caused, besides other reasons, by worse accessibility to the centres, i.e. by parameters of the transport network. The transportation factor is thus manifested relatively significantly within the competition between individual communities. However, with regard to the intraregional aspect, the rural areas represent a relatively homogenous area.

By comparing particular time periods (the original survey started already in 1869, when the first population census was executed on the Habsburgs’ Austrian Empire; see Drápela, 2010) we may conclude that both Figures differ mainly by the intensity of continuing depopulation. This is caused by demographical trends going on within the Czech population, where the period between 1991 and 2001 was a period of historically lowest population growths. On the other hand, the period between 2001 and 2010 was a time period when the population of strong age groups of the 70s of the 20th century were starting to establish families. What is important though, is that the spatial allocation of depopulation areas has not changed significantly, which can be said even after comparing data older than 100 years (see Drápela, 2010). The problems of these regions are apparently of a long term character, which is an important finding.
CORRELATION ANALYSIS

The complex character of marginalization, manifested by depopulation, results in the impossibility to definitely mark a factor or a group of factors as a cause of the phenomenon. Various factors are mutually interlaced and it is frequently impossible to separate the effects of one factor from another. For example the land relief and the watercourse network in upland areas often affects the centrality of such location, the leading of main roads, fertility of soil and alike. Economic, social and demographic factors are similarly interwoven, too. That is why the analysis is focused on revealing the typical features of depopulation areas and it does not judge, for the time being, what is the cause and what the consequence. The research method applied was a classic correlation and regression. With regard to the character of the phenomenon and to the fact that only the similarity of division is studied and not the mutual dependence of variables, indicators with a Pearson correlation coefficient value of at least ±0.3 were assessed as significantly similar. The following 17 indicators (in brackets their abbreviations used in Table 1) have been selected for the analysis:

- the average age of community inhabitants (AverAge)
- proportion of inhabitants aged 0 – 14 of the total number of inhabitants (Inhab_0-14)
- proportion of inhabitants aged 65 and more of the total number of inhabitants (Inhab_65+)
- proportion of inhabitants aged 15 and more with a completed university degree of the total number of inhabitants (Prop_UNI)
- proportion of inhabitants aged 15 and more with a completed university degree, higher technical school or secondary school with maturita (school leaving exam) of the total number of inhabitants aged 15 and more (UNI.HTS.SSm)
- proportion of inhabitants aged 15 and more with elementary or uncompleted elementary education, or with no education of the total number of inhabitants aged 15 and more (EL+without)
- number of inhabitants (InhabNum)
- proportion of persons economically active of the total number of inhabitants (Prop_EA)
- proportion of employed persons of the total number of economically active persons (PropEmp)
- proportion of unemployed persons of the total number of economically active persons (PropUnemp)
- proportion of employers and self-employed persons of the total number of economically active persons (Prop_SelfEmp)
- proportion of economically active persons working in primary sector of the total number of economically active persons (Primar)
- proportion of economically active persons working in tertiary sector of the total number of economically active persons (Tertiar)
- proportion of persons commuting daily to work outside the community of the total number of employed persons (ComDaily)
- proportion of persons commuting out of the district of the total number of employed persons (ComOutDis)
- proportion of persons whose commuting time is 30 min and more of the total number of employed persons (ComTime30+)
- proportion of persons whose commuting time is 60 min and more of the total number of employed persons (ComTime60+)

The results of correlation analysis for individual indicators are shown in Table 1.

Table 1: The values of Pearson correlation coefficients for used indicators

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>AverAge</td>
<td>-0.43383</td>
</tr>
<tr>
<td>Inhab_65+</td>
<td>-0.40756</td>
</tr>
<tr>
<td>Primar</td>
<td>-0.35877</td>
</tr>
<tr>
<td>EL+without</td>
<td>-0.30840</td>
</tr>
<tr>
<td>PropUnemp</td>
<td>-0.01534</td>
</tr>
</tbody>
</table>
Table 1 implies that the most significant features of depopulation are the age structure of inhabitants, branch structure of economy in these areas and education level of inhabitants. The age structure of inhabitants was represented by the following indicators: AverAge (correlation coefficient value -0.43), Inhab_65+ (-0.41) and Inhab_0-14 (0.38). The Pearson correlation coefficient values imply that in the Czech Republic depopulation areas the average age of inhabitants is much higher than in other areas, as well as the proportion of inhabitants in post-productive age, whereas the proportion of inhabitants in pre-productive age is significantly lower here. This fact, however, also arises from the characteristics of the inhabitant mobility indicator, composed of mechanical as well as natural inhabitant mobility.

The group of indicators reflecting the branch composition of economy is represented by the following indicators: Tertiar (0.37) and Primar (-0.36). Depopulation areas are characterized by a high proportion of employed persons working in agriculture, forestry and fishing and to the contrary, by a low proportion of employed inhabitants working in various branches of services.

The last significantly manifesting group of factors is the education level of inhabitants, represented by the following factors: Prop_UNI (0.33), EL+without (-0.31) and UNI.HTS.SSm (0.29). In depopulation areas there is a lower proportion of inhabitants with higher education levels (especially inhabitants with a completed university degree) and to the contrary, a higher proportion of inhabitants with lower levels of education or even without any.

We could consider the values of indicators Prop_EA, Prop_Emp and Prop_SelfEmp as showing weak similarities, while the remaining indicators reach values too near to zero to be considered as manifested in any way in depopulation. Somewhat surprisingly the transportation indicators related to commuting are among them. The ComOutDis indicator shows the highest correlation coefficient values, which is not surprising as these areas are often situated near administrative boundaries. Commuting behind the boundaries is thus not more time consuming than commuting to a community within the territory of their own district. The time taken by commuting to work is also practically the same as in the areas with
a growing number of inhabitants. It can be said that the parameters of commuting from depopulation areas do not differ in any way from the rest of the country. On the contrary, it can be assumed that depopulation areas are less attractive for their inhabitants because of low well paid job opportunities in perspective branches (see Tertiar, Primar), which causes the drift of educated people to centres (see Prop_EDU, ELZS+without) and with regard to the fact that the more educated are mostly young people, the population of such areas gets older (see AverAge, Inhab_65+) and accordingly, depopulated.

RESULTS OF CASE STUDIES

To improve understanding of transport impact on the process of area depopulation and to find factors which are unidentifiable by the quantitative analysis, case studies were executed for three model areas, which in maximum extent focus on factors that influenced the current central structure in the past and that were the causes of the depopulation in these areas. The questionnaire surveys show that the most important factors that make life difficult for inhabitants of depopulated areas are poor road condition, insufficient number of public transport services, financially demanding transport, and poor road maintenance in the winter season. These factors were in the questionnaires marked as problematic by more than 30% respondents, and even in open questions, most objections regarded those factors. Their importance grows with smaller size of a municipality or its part where the respondent lives. Small municipalities and their parts are often only connected by poor quality local roads, which are considerably worse maintained than main roads and which are often very old. Unless the villages are located on major routes, public transport services usually connect small villages less frequently, and the public transport stops are often located only on junctions of main roads and the local roads connecting the villages.
Particularly the problematic funding of public transport in depopulating areas make the inhabitants depend on their passenger cars, since the public transport is unable to meet all their needs. However, the use of passenger cars is a relatively expensive matter for individuals, especially when taking into account significantly lower average salaries in rural areas than in towns. Consequently, in some cases, improvised forms of car sharing appear in rural areas, and according to the respondents it is possible to decrease costs even lower than costs for public transport.

The annoying poor road conditions for respondents, which is caused by their age and insufficient maintenance, together with the non-existence of roads to some local parts (particularly groups of solitary houses), is subsequently reflected in the development of population number in such remote localities. In some extreme cases, small villages were completely abandoned by permanent residents and the existing housing is only used for recreation purposes. However, this process is not irreversible. In some localities where new roads were built and where the aesthetic value of the surrounding countryside is high, the houses became used again for permanent residency. In less attractive areas, particularly production agricultural or industrial areas, this positive trend is only hardly expected, since the abandoned buildings are unused even for recreational purposes and gradually dilapidate.

Interesting results were found within a historical-geographical analysis in the issue of transport infrastructure importance for towns and villages development, when unambiguously the higher quality transport infrastructure has significantly positive impact on development opportunities in a region. In model areas some cases were documented, when a construction
of an important road enhanced the significance of a municipality located on such road, in comparison with other municipalities located off that road. The most well-known example in the Czech Republic is the decline in importance of the former regional centre town of Chrudim. It resisted the connection to a railway at the expense of the current regional centre town of Pardubice, which is now located on the busiest railway in the Czech Republic. This is the reason why Pardubice significantly outgrew its competitor. However, the importance of the location on a main road decreases with the size of a municipality. Small villages have only little benefit from such roads, unless they can offer services. The importance of main roads does not decrease in mountainous regions, where the passability of local roads is worse.

The research of spatial mobility of population shows, that localities most affected by depopulation are such localities where respondents spend minimum of their free time outside their home. If a given municipality is unable to offer facilities for spending free time and if such municipality is unattractive for residing from a certain reason, there is a gradual migration from such a place to more convenient localities. This unattractiveness of some regions is reflected in spatio-temporal travel behaviour of the population, which in turn leads to a centralization of some larger municipalities where inhabitants of small village spend more time.

Another important finding is the sensitivity of rural areas inhabitants to road fee collection, which was discovered in the model area of Pelhřimov region. Despite the fact that this area is directly adjacent to the artery of the whole Czech Republic, motorway D1, the inhabitants use it rarely for commuting, unless they travel longer distances. A large number of inhabitants of this model area commute to the town of Humpolec, which is located directly on the motorway. Most respondents chooses a route along local roads, since according to their answers, they are unwilling to be charged for using the motorway while there is an option just several minutes longer. Therefore, the inhabitants do not use the potential of the motorway, since it is too expensive for them.

**DISCUSSION**

If the obtained results are compared with other studies dealing with rural area marginalization problem, we find out they come to similar conclusions, though using different methods. Definition of problem areas, for example, corresponds very much with the study of Musil and Müller (2006), who had analysed a range of indicators from economic and social sphere. Nevertheless, the definition on the basis of totally different indicators is, to a large extend, in accordance with the definition of problem areas on the basis of inhabitants mobility demographic indicator. Many similarities can be analogically found with studies of Havlíček and Chromý (2001), Řehák (2004), Čermák (2005) and others. The time dimension of this hard to solve problem can be understood when the current results are compared with older studies, e.g. Musila (1988) or Řeháka (1979). The geographical allocation of marginalized areas has not been changed for long decades, which raises the question whether it is caused
only by wrong dealing with the problem, or whether these areas cannot be successfully
developed just because of the current state of society, life style and technology solutions.
The results of the quantitative analysis can be compared with fewer studies than in the
previous case, where the most relevant is the study by Marada (2001). He focuses in his study
especially on physio-geographic and administrative factors, where from the analyses arise the
fact that the agriculture areas, more extensively exploited, are often marginalized areas. By
comparison with state administrative boundaries of various levels Marada comes to the
conclusion that most boundaries have no effect on the allocation of marginalized areas, but it
is rather the factor of main centre distance that matters. This finding is in accordance with the
results of the Commuting Out of the District Boundaries Indicator. The fact that the distance
of commuting has not a big effect on depopulation reduces meaningfulness of some studies
working with this factor (e.g. Rehák, 1979).
The case studies carried out represent rather isolated research in the Czech Republic focused
on problematic rural areas, where such a specialized sociological research has been carried
out here only once, within the study of border areas (Jeřábek, Dokoupil, Havlíček, et al.,
2004). The specific focus of the survey on transportation and spatial mobility of inhabitants is
unique within the Czech specialized literature. No survey of travel behaviour of inhabitants in
rural areas had ever been executed here before this study and so it is very difficult to compare.
Some of the rural areas problems are more often reflected by media than by specialized
literature (e.g. the issue of rural roads maintenance), where the professional public rely on
hard data rather than on opinion polls. It is apparent that even the data on the number of
kilometres of repaired roads can be misleading, as the roads in the worst condition are not
repaired at all. On the contrary, the opinions on public transport frequency are always to be
taken with a pinch of salt, as demand for it will always be higher than offer. However, the
finding how significantly the financial factor affects transport behaviour of inhabitants (not
using motorways due to money saving, as the year motorway toll costs about 50 EUR) is
certainly applicable to be taken into account.

RECOMMENDATIONS

As mentioned above, the issue of depopulation of some areas is largely caused by their
distance from strong centres. The distribution of these centres in time is relatively stable and
therefore the distribution of problematic areas is also a long-term matter. The options to
reverse the negative trend and development of these areas are rather limited, since the
reduction of the spatial distance factor may only be reached with difficulties. The only way to
face this factor is virtually just the technological progress together with the development of
transport infrastructure in such localities, which allows for lower commuting times to work.
However, this strategy faces the fact that in rural areas the population is less educated and
professionally focused on work in agriculture and industry. Therefore, the population of rural
areas is unable to compete with better qualified inhabitants of more advanced regions, thus the necessary transport costs will still be a burden, which will keep them demotivated when taking into account the low salaries in not very prosperous industries. The solution may lie in supporting local centres. In the future, local centres may play a more important role in developing depopulated regions, since they may have jobs corresponding with the education and qualifications of inhabitants of these remote areas. In order to ensure the prosperity of these centres and a sufficient number of jobs for their inhabitants, it is essential to provide good quality infrastructure to connect higher level centres. Not only transport infrastructure is considered, but also the technological one, while currently it is the high-speed internet which has the highest priority. It is probable that despite strengthening local centres, some smaller villages will still face complete depopulation. However, it is a process which has always been present. Regarding the technology, it is necessary to focus on inexpensive and energetically economical technologies, since the financial factor significantly affects decisions of inhabitants in problematic areas. What may help is more extensive service of some transport modes such as car sharing a car pooling, while it would be necessary to adapt these ideas to specific target groups. Similarly, new approaches and optimization of public transport organization may improve the mobility of population of these areas, or reduce financial factors of commuting to work. And last but not least, it is essential to provide good maintenance of the existing roads, since building new roads is unnecessary, if the existing ones are of good quality.

CONCLUSION

Fighting the decline of remote rural regions is one of the priorities of the European regional policy. Due to the complex nature of causes of this issue, finding a solution is a difficult task which requires measures in different areas. This article aimed to explain which measures may be applied in the field of transport, in order to limit the negative factors as much as possible, or even reverse them. The results of the present study show that merely the presence of transport infrastructure is insufficient to stimulate regional development; it is crucial to take into account the qualitative conditions of the infrastructure as well as the parameters which are perceived by its users. Inhabitants of depopulated areas are characterized by lower education and qualification in worse paid jobs, which needs to be taken into account when designing measures in order to support their mobility. These measures should be very sensitive to economic demands of transport. Although building new good quality roads will not always bring about the intended effects on regional development, it is to say that in general it is beneficial with regard to development. Considering the financial demands when building new infrastructure, local centres need to be found, based on their commuting conditions, and connected to quality infrastructure that would allow for delivering products which are produced there to
consumers. The local centres would then may become development poles and prevent depopulation and decline of the existing marginalized regions.

REFERENCES


ELASTICITY OF LONG DISTANCE TRAVELLING

ABSTRACT

With data from the Danish expenditure survey for 12 years 1996 through 2007, this study analyses household expenditures for long distance travelling. Household expenditures are examined at two levels of aggregation having the general expenditures on transportation and leisure relative to five other aggregated commodities at the highest level, and the specific expenditures on plane tickets and travel packages at the lowest level. The Almost Ideal Demand System is applied to determine the relationship between expenditures on transportation and leisure and all other purchased non-durables within a household. Due to a high share of corner solutions among the expenditures on plane tickets and package travelling, the expenditures on these specific commodities are examined with a Tobit approach. The model results find both plane tickets and travel packages to be luxury goods. It also states that travel packages has higher income elasticity of demand than plane tickets but also higher than transportation and leisure in general. The findings within price sensitivity are not as sufficient estimated, but the model results indicate that travel packages is far more price elastic than plane tickets which are actually found close to be unit elastic.

1 INTRODUCTION

Within the transportation research field, the relevance of long distance travelling has increased as transport models in general have increased in proportions as well as in details. Generally, most focus is put on daily transportation even though the relative limited number of outbound travel represent approximate 40 percent of the total travelled kilometres (Knudsen, 2010). Apart from composing a significant share of the total amount of travelling, holiday travel is most likely sensitive to economic changes and hence very relevant when forecasting the future amount of travelling. It is of high relevance to improve the knowledge of the interactions between the
amount of travelling and general economic changes in society in order to forecast long distance travelling.

Due to the available data, private expenditures on plane tickets and travel packages will act as a mean to analyse consumption on long distance travel even though several other types of travelling also seems relevant. These are however not isolated described within this data.

The paper is structured as follows. In the next section, the formulation of the two applied models are outlined and in Section 3 the survey is described including an overall descriptive analysis of the commodities of most interest. In Section 4 the final model specifications are listed and Section 5 presents the results of the estimations. Finally in section 6 the findings are summarised with some concluding remarks.

2 MODELLING FRAMEWORK

In order to analyse the expenditure shares of aggregated commodities the Almost Ideal Demand System (AIDS, Deaton and Muellbauer, 1980) is applied. AIDS is one of the most widely used approaches for estimating empirical demand systems in consumer demand analysis (see e.g. Chern, 2003; Hausman, 1997; Chang, 2010). The model system is an extension of the Working-Leser model (Working, 1943; Leser, 1963) in which the budget share of good \( i \) is linearly related to the logarithm of prices and total real expenditures. This system approach has an advantage over the single equation approach in estimating empirical demand systems as it can analyse the interdependence of budget allocations for different consumer goods and services.

The expenditure share \( w_i \) associated with the \( i \)-th good provides the general form of the AIDS model (1):

\[
w_i = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log \left( \frac{X}{P} \right)
\]

where \( p_j = (p_1, \ldots, p_J) \) is the price vector of all goods included in the model, \( X \) is the total household expenditure and \( P \) is a nonlinear price index given by (2). \( \alpha \), \( \beta \) and \( \gamma \) are the parameters to be estimated.

\[
\log P = \alpha_0 + \sum_i \alpha_i \log p_i + \frac{1}{2} \sum_i \sum_j \gamma_{ij} \log p_i \log p_j
\]

where \( \gamma_{ij} = \gamma_{ji} \)
Due to well-known properties of an expenditure function (see, e.g., chapter 3 in Mas-Colell, 1995) the parameters $\gamma_{ij}$, $\alpha_i$ and $\beta_i$ must satisfy the restrictions listed in (3) and (4):

\begin{align}
\text{homogeneity} & \quad \sum_j \gamma_{ij} = 0 \\
\text{adding up} & \quad \sum_{i=1}^n \alpha_i = 1 \quad \sum_{i=1}^n \gamma_{ij} = \sum_{i=1}^n \beta_i = 0
\end{align}

Changes in relative prices work through the $\gamma_{ij}$ parameters and changes in expenditures operate through the $\beta_i$ parameters. The adding up restrictions ensure that the sum of $\beta_i$ adds up to zero. In the model, $\beta_i$ being positive means that the income elasticity is above 1, whereas $\beta_i$ being negative means elasticity below unity. In this way, positive values of $\beta_i$ correspond to luxury goods whereas negative values correspond to necessities (Deaton and Muellbauer, 1980).

The underlying theory of AIDS assumes no corner solution among individuals, i.e. each household is assumed to have expenditures on all commodities. This implies that the model system is mainly appropriate for aggregated expenditure data. When looking at a detailed expenditure portfolio within a limited time frame, zero consumption may occur. In that case, additional conditions or other approaches need to be considered (see, e.g., Brännlund and Nordström, 2002; Chern et al., 2003). Even so, the AIDS approach has been widely applied on more disaggregated goods even as specific as brands (see, e.g., Hausman, 1997) by estimating AIDS to analyse the demand for cereal products.

The most commonly applied models to account for zero expenditure shares are Heckman’s two step model (Heckman, 1979) and the standard Tobit model (Tobin, 1958; Ameiyda, 1974). Each model is based on different assumptions regarding zero consumption.

If zero consumption is assumed to be due to sample selection, in the sense that no purchase of the particular item was made during the survey period (e.g., because of a short survey period), Heckman’s two-step model is the appropriate model. The Tobit model on the other hand simply captures the corner solutions for utility maximisation, where zero actually represents no expenditures on the specific good (Brännlund and Nordström, 2002).

In the present expenditure survey to be described in Section 3, zero consumption might easily occur as a combination of both sample selection and “true” zero consumption for certain household types. For a significant share of the included commodities, data are collected within a two weeks period, and hence the probability of having the corner solution representing infrequency of purchase is high. However, some specific and rare commodities (especially
durables) are also likely to actually have zero consumption within some households during a particular period of observation. Chern et al. (2003) handle this difference in explaining corner solution by applying both the Heckman and the Tobit approach to outline a probable interval span of the elasticities. Due to the purchase of plane tickets and travel packages, it is reasonable to assume that the corner solutions actually correspond to zero expenditures and this is the reason why the Tobit model is applied.

The Tobit approach is an econometric model for censored endogenous variables proposed by James Tobin (Tobin, 1958). It was developed to describe the linearly relationship between the non-negative dependent variable $w_i$, and a linear predictor $\beta z_i$. The Tobit model is:

$$ w_i = \begin{cases} w_i^* & \text{if } w_i^* > 0 \\ 0 & \text{if } w_i^* \leq 0 \end{cases} $$  \hspace{1cm} (5)

where $w_i^*$ is the latent variable: $w_i^* = \beta z_i + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$.

The specification of $w_i^*$ in this study is based on the model specification of the expenditure shares used in AIDS though without applying the general price index $P$, cf. (6)$^1$:

$$ w_i^* = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log X + \epsilon_i $$  \hspace{1cm} (6)

3 Data

The modelling effort in the present paper is based on the Danish expenditure survey. The survey registers the total expenditure of Danish households divided into approximately 1,300 commodities. Data cover a 12-year period from 1996 to 2007 and include expenditure information from approximately 900 households per year extracted as a representative sample of the population. For all households, detailed background information is available, including among other things information about composition and income.

The survey only registers total expenditures on the different commodities and does not provide any information on unit prices or product quantities. Instead, prices from the national Price Index

$^1$ The stochastic process of the Tobit model determines in this case expenditure shares and the discrete switch at zero.
have been used (Statistics Denmark, 2011). These prices are divided into similar classifications of commodity groups as applied in the model.

Table 1 below shows some properties of the six aggregated commodity groups applied in the model. Plane and travel packages are included in the relative broad defined commodity group of transportation and leisure which accounts for 43 percent of the total expenditures. All expenditures are corrected for inflation and possible outliers are removed from the survey.

<table>
<thead>
<tr>
<th>Commodity Group</th>
<th>Obs.&gt;0</th>
<th>Share of zero cells</th>
<th>Average Expenditure share</th>
<th>Average yearly change in expenditure shares</th>
<th>Average yearly change in prices</th>
<th>Average yearly change in total quantities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food, drinks and tobacco</td>
<td>9,538</td>
<td>0 %</td>
<td>25 %</td>
<td>-2.2 %</td>
<td>1.6 %</td>
<td>0.9%</td>
</tr>
<tr>
<td>Clothes and footwear</td>
<td>7,870</td>
<td>17 %</td>
<td>7 %</td>
<td>-0.7 %</td>
<td>-1.4 %</td>
<td>2.6%</td>
</tr>
<tr>
<td>Electricity and heating</td>
<td>9,507</td>
<td>0 %</td>
<td>15 %</td>
<td>1.3 %</td>
<td>3.8 %</td>
<td>4.5%</td>
</tr>
<tr>
<td>Medicine and medical care</td>
<td>8,699</td>
<td>9 %</td>
<td>3 %</td>
<td>2.4 %</td>
<td>1.3 %</td>
<td>6.4%</td>
</tr>
<tr>
<td>Communication and audio equipment</td>
<td>9,474</td>
<td>1 %</td>
<td>7 %</td>
<td>0.5 %</td>
<td>-4.3 %</td>
<td>3.6%</td>
</tr>
<tr>
<td>Transportation and leisure</td>
<td>9,538</td>
<td>0 %</td>
<td>43 %</td>
<td>0.6 %</td>
<td>2.6 %</td>
<td>4.1%</td>
</tr>
</tbody>
</table>

The four top commodity groups correspond to the overall grouping from the expenditure survey, whereas the final two commodity groups are rearrangements of the original commodity groups to have plane and travel packages within the same group. Besides long distance travel, this commodity group contains very different commodities such as toiletries, daily transportation, expenditures on pets, education etc. The price index applied is estimated as a weighted average from market shares found in the expenditure survey and is hence biased.

The total expenditures applied in the model are the total expenditures on the non-durables listed in Table 1. The durables include housing expenditures, house equipment, purchase of cars and other means of transport, while large consumer durables for leisure activities are excluded from the survey. Since the durables are excluded from the survey, the model specification implies that there is no substitution between durables and non-durables.

During the 12-year survey period, the expenditure share of food, drinks and tobacco decreases whereas the total quantities purchased of drinks and tobacco are relatively stable. This implies a general increase in disposable income resulting in increased consumption on other goods and services. The stated quantities in the table are total quantities describing the total unit purchased within the commodity group derived from total expenditures divided by the price index. Hence,
this figure does not describe changes in actual quantities purchased or changes in quality of the goods.

Generally, the total quantity increases within each commodity group, but the yearly increase of 4.1 percent within transportation and leisure compared with the high expenditure share indicates that quite a substantial share of the extra disposal income is spent for transportation and leisure.

### 3.1 Long distance travel

The commodities having most interest are the plane tickets and travel packages. These two commodities act as instruments for the analysis of long distance travel even though they do not describe all outbound travel.

Figure 1 and Figure 2 show some properties of the two commodities. Figure 1 shows the development in the prices of travelling from 1996 to 2007 compared with the general price index. Figure 2 shows the changes in expenditure shares of the two commodities and an aggregated travel commodity.

The prices of plane tickets and travel packages have increased by 1.6 and 3.1 percent per year, respectively. The prices of travel packages have increased somewhat more than the 1.9 percent yearly change in the total price index. The changes in expenditure shares are higher than the changes in prices. In total, travelling have increased by 5.4 percent per year, whereas plane tickets have increased by 8.7 percent and travel packages only by 4.5 percent. This very high growth in budget shares should be considered relative to the very small average expenditure shares of 0.5 and 1.8 percent, respectively.

The total expenditures on most of the commodities are registered within a two weeks period, but the consumption on more infrequent commodities is registered within a year. That is the case for
plane travel and travel packages ensuring a higher sample. Still, only 13 and 28 percent of the households register expenditures on plane tickets and travel packages, respectively. Combining the two commodities to an overall travel commodity, only 38 percent of the households register expenditure on travelling.

This high share of non travellers does not reflect the general share of non travellers within the population since for example car travelling is not included. From statistic Denmark’s registration of “Business and holiday travel 2007” (Statistic Denmark, 2008), it appears that approximately 40 percent of the population have holidays with plane as primary mode. Due to this it seems reasonable to assume that the high number of corner solutions is probably not due to sample selection which justifies the Tobit approach over Heckman’s.

4 MODEL SPECIFICATION

The model estimation is separated into two parts: firstly, the overall elasticities are estimated with the AIDS model for the six commodity groups described in Section 3; secondly, long distance travel is analysed in more detail with the Tobit model.

The expenditure functions applied in both cases are similar linear relations describing the observable expenditure shares as a function of the logarithm of prices and the logarithm of total expenditures. The AIDS model is derived from the Working-Leser model and to include heterogeneity in population, six household specific constants are applied to the model specification as in (7):

$$w_i = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log(X/P) + \sum_k \delta_i H_k$$

The applied model specification in the Tobit model is also based on the Working-Leser specification and the same household specific constants are applied to include heterogeneity:

$$w_i = \begin{cases} w_i^* & \text{if } w_i^* > 0 \\ 0 & \text{if } w_i^* \leq 0 \end{cases}$$

where

$$w_i^* = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log X + \sum_k \delta_i H_k + \varepsilon_i$$

The vector $H_k$ includes the six household specific constants, where the household members are grouped into five age groups: babies up till two year olds, small children before starting in school (3-6), children before the teens (7-11), teens (12-17) and all household members above 17 are grouped as grownups. The final variable is the number of retired persons in the household.
Furthermore, a dummy variable representing singles was tested but with very limited impact. A
time trend dummy was also implemented, but it was found to be insignificant.

5 Results

The results are divided into two parts reflecting the two applied approaches; the results of the
aggregated AIDS model describing the overall relations between commodities, and the
disaggregated Tobit mode describing the demand for travelling.

5.1 Aggregated AIDS estimation

Table 2 shows the parameter estimates from the AIDS approach as described in Section 2. Due to
the level of aggregation in the commodity groups, the problem of corner solutions can be
ignored.

**TABLE 2 - PARAMETER ESTIMATES OF THE AGGREGATED AIDS MODEL**

<table>
<thead>
<tr>
<th></th>
<th>Food, drinks and tobacco</th>
<th>Clothes and footwear</th>
<th>Electricity and heating</th>
<th>Medicine and medical care</th>
<th>Communication and audio equipment</th>
<th>Transportation and leisure</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{Food,drinks and tobacco}$</td>
<td>-0.0178</td>
<td>0.0570**</td>
<td>0.0630*</td>
<td>0.0550**</td>
<td>0.00923</td>
<td>-0.166**</td>
</tr>
<tr>
<td>$\gamma_{Clothes and footwear}$</td>
<td>0.0570**</td>
<td>-0.0494</td>
<td>-0.116**</td>
<td>0.00491</td>
<td>0.0292**</td>
<td>0.0748*</td>
</tr>
<tr>
<td>$\gamma_{Electricity and heating}$</td>
<td>0.0630*</td>
<td>-0.116**</td>
<td>-0.0277</td>
<td>-0.0251</td>
<td>0.0160</td>
<td>0.0902</td>
</tr>
<tr>
<td>$\gamma_{Medical and medical care}$</td>
<td>0.0550**</td>
<td>0.00491</td>
<td>-0.0251</td>
<td>-0.0190</td>
<td>-0.0189*</td>
<td>0.00301</td>
</tr>
<tr>
<td>$\gamma_{Communication and audio}$</td>
<td>0.00923</td>
<td>0.0292**</td>
<td>0.0160</td>
<td>-0.0189*</td>
<td>-0.0318*</td>
<td>-0.00377</td>
</tr>
<tr>
<td>$\gamma_{Transportation and leisure}$</td>
<td>-0.166**</td>
<td>0.0748*</td>
<td>0.0902</td>
<td>0.00301</td>
<td>-0.00377</td>
<td>0.002294</td>
</tr>
</tbody>
</table>

| $\beta_l$          | -0.0760**               | 0.0410**             | -0.0540**               | 0.00483**                  | -0.0151**                        | 0.0994**                   |
| $\alpha_l$         | 0.548**                 | -0.118**             | 0.412**                 | 0.00974**                  | 0.165**                          | -0.0165                    |
| **Babies (0-2 years)** | 0.00492                 | 0.00544*             | 0.00242                 | -                          | -0.0102**                        | 0.00115                   |
| **Small children (3-6 years)**       | 0.00960**               | 0.000923             | 0.00327                 | -                          | -0.00783**                       | -0.00219                  |
| **Children (7-11 years)**            | 0.0176**                | -0.00206             | 0.00684**               | -                          | -0.00316**                       | -0.0139**                 |
| **Teens (12-17 years)**              | 0.0240**                | 0.00452**            | 0.00705**               | -                          | 0.00428**                        | -0.0373**                 |
| **Grownups (above 17 years)**        | 0.0372**                | -0.0179**            | 0.00307**               | -                          | -0.00331**                       | -0.0165**                 |
| **Retired persons**                  | 0.00492                 | -0.00628**           | 0.0251**                | -                          | -0.0162**                        | -0.0287**                 |

** and * indicate that estimates are significantly different from zero at the 0.05 and the 0.10 level respectively

The estimation of the price related parameters suffers from lack of variability in prices. As
aforementioned in Section 3, prices are measured on the basis of average price index for the
commodity group (over time) and are not specific to individuals. However, the model estimation appears reliable and the remaining model parameters are generally significant with a high level of confidence.

All $\beta$ parameters are estimated with high confidence and provide a good basis for the estimation of income elasticities. Obviously this is due to the fact that income is household specific in contrary to prices and hence includes much more variation. Income elasticities can be measured as:

$$e_i = \frac{\beta_i}{w_i} + 1$$

(9)

Table 2 shows the estimated income elasticities and for comparison also a number of estimates from the USDA-Economic Research Service calculation using 2005 ICP data (ICP, 2005). All the commodity groups applied in the two sets of estimations are not completely identical, but the food, clothes and medicine commodities are apparently the same.

<table>
<thead>
<tr>
<th>TABLE 3 – INCOME ELASTICITY ESTIMATED FROM THE RESULTS OF THE AIDS MODEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food, drinks and tobacco</td>
</tr>
<tr>
<td>Income elasticity</td>
</tr>
<tr>
<td>USDA</td>
</tr>
</tbody>
</table>

The aggregated AIDS model finds the three commodity groups: food, drinks and tobacco, electricity and heating, and communication and audio equipment to be necessities. Clothes and footwear, medicine and medical care, and transportation and leisure are estimated to be luxury goods. In the USDA estimations, the income elasticity of clothes and footwear are found to be approximately one. This difference might be due to a significant share of corner solutions, which will tend to upward bias the corresponding elasticity in the AIDS model.

5.2 Tobit estimation of plane tickets and travel packages

As mentioned in Section 3, only 39 percent of the households have registered consumption on plane tickets or travel packages during the year they participated in the survey. Due to this, applying an AIDS on these disaggregated commodities is theoretically unsuitable. Instead, the Tobit model described in Section 2 is applied.
As were the case for the aggregated AIDS estimation, the estimation of the Tobit model also suffers from the lack of price variation in the data. The Tobit model is modelled as a multivariate model system of the budget shares of plane tickets and travel packages relative to the total expenditures on travelling.

Table 4 shows the parameter estimates of the model system. Most parameters are significant and, as for the aggregated AIDS model, the $\beta$ parameters are estimated with high confidence and thereby give a good basis for the estimation of income elasticities. As can be seen, the own price related parameter for plane tickets has not been significant estimated. But the standard error is though too high assume the estimate equal zero.

Obviously, the lack of price variation is a problem for the identification and interpretation of price elasticities. However, it does not prevent the estimation of income elasticities that are correct on average as long as the separability of the Tobit equation in terms of $X$ and $p$ holds.

As the model estimates are the result of a Tobit approach, the interpretation of the parameters is not completely parallel with the traditional interpretation of the parameters. These parameters also represent the latent choice of actually purchasing the specific good.

The household specific constants included in the model in general show that an increasing number of household members have negative impact on the expenditure shares of plane tickets and travel packages. The number of retired household members has the highest impact on the purchase of plane tickets, whereas the existence of babies in the household has the highest impact on travel packages.

The significant $\rho$ indicates existence of correlation between the purchase of plane tickets and travel packages. This justifies the multivariate model system instead of two separate univariate Tobit models.
Where unconditional elasticity represents the whole population hence also includes the households having zero expenditure, the conditional elasticity reflect the econometric relations for travellers rather than the whole population. The conditional elasticities are estimated from the relations in (10), (11) and (12).

$$e_i = \frac{\partial E[q_i|q_i^* > 0]}{\partial X} \frac{X}{E[q_i|q_i^* > 0]} = 1 + \frac{\hat{\alpha}_i \left[ 1 - \hat{z}_i \frac{\phi(\hat{z}_i)}{\Phi(\hat{z}_i)} - \left( \frac{\phi(\hat{z}_i)}{\Phi(\hat{z}_i)} \right)^2 \right]}{\bar{x}_i \hat{\beta} + \hat{\sigma}_i \frac{\phi(\hat{z}_i)}{\Phi(\hat{z}_i)}} \tag{10}$$

$$e_{ii} = -1 + \frac{\hat{\beta}_{ii} \left[ 1 - \hat{z}_i \frac{\phi(\hat{z}_i)}{\Phi(\hat{z}_i)} - \left( \frac{\phi(\hat{z}_i)}{\Phi(\hat{z}_i)} \right)^2 \right]}{\bar{x}_i \hat{\beta} + \hat{\sigma}_i \frac{\phi(\hat{z}_i)}{\Phi(\hat{z}_i)}} \tag{11}$$

$$e_{ij} = \frac{\hat{\beta}_{ij} \left[ 1 - \hat{z}_i \frac{\phi(\hat{z}_i)}{\Phi(\hat{z}_i)} - \left( \frac{\phi(\hat{z}_i)}{\Phi(\hat{z}_i)} \right)^2 \right]}{\bar{x}_i \hat{\beta} + \hat{\sigma}_i \frac{\phi(\hat{z}_i)}{\Phi(\hat{z}_i)}} \tag{12}$$

where $\hat{z}_i = \frac{\bar{x}_i \hat{\beta}}{\sigma}$

The estimated elasticities are listed in Table 5 together with the elasticities estimated from two model approaches applying restrictions on the parameters.

<table>
<thead>
<tr>
<th></th>
<th>No restrictions</th>
<th>Restrictions on $\gamma$</th>
<th>All restrictions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Income Elasticity</td>
<td>Price elasticity</td>
<td>Income Elasticity</td>
</tr>
<tr>
<td></td>
<td>Plane Package</td>
<td>Plane Package</td>
<td>Plane Package</td>
</tr>
<tr>
<td>Plane tickets</td>
<td>1.12 -0.73 0.33</td>
<td>1.12 -1.57 0.57</td>
<td>0.93 -1.16 0.16</td>
</tr>
<tr>
<td>Travel packages</td>
<td>1.36 -0.33 -1.80</td>
<td>1.35 1.06 -2.06</td>
<td>1.41 0.97 -1.97</td>
</tr>
</tbody>
</table>

The income elasticity of travel packages is higher than the one of plane tickets. This relation is observed for all the different applied model specifications. The estimates of 1.1 and 1.4 respective are in line with the findings of the commodity group of transportation and leisure in the aggregated AIDS model.
The conditional own price elasticities state that the purchased quantities of travel packages are more sensitive to price changes than is the case for plane travel. This has actually also turned out to be consistent across the different applied model specifications despite the uncertainty of the parameter estimates. In contrary, the different applied approaches have not provided similar consistent trends concerning the cross price elasticities. In this multivariate model system, the demand for travel packages decreases when the prices of plane tickets increase, whereas the demand for plane tickets increases when the prices of travel packages increase. This actually seems intuitive as travel packages most often also includes plane travelling.

The elasticities are further estimated per year. Generally, the yearly change is not great in magnitude, but during the 12-year period the income elasticity of plane tickets has increased by 1.3 percent and the income elasticity of travel packages has decreased by 2.0 percent. This indicates a change in the perception of the two travel types during the survey period. But while the income elasticity increases continuously during the years, the change within travel packages is observed from 1999 to 2002 and hereafter the income elasticity seems relatively constant. This obvious change within the income elasticity of travel packages corresponds to the observed change in the expenditure share of plane tickets observed in Figure 2 in Section 3.1 as well as the increase in the price of travel packages during a period of relative constant plane ticket prices as illustrated in Figure 2.

Even though the changes in income elasticities have opposite directions, the income elasticity of travel packages is still 19 percent higher than the income elasticity of plane travel in 2007. From these limited changes per year it is not certain that the two income elasticities will intersect in the future.

Due to the uncertainties related to the estimation of $\gamma$, two different groups of restrictions are added to the model specification to improve the estimates as listed in Table 5. Both restrictions are similar to the restrictions applied to the AIDS model in (3) and (4): the first group of restrictions is applying the restrictions on $\gamma$, whereas the second is applying all the restrictions from the AIDS model.

The applied restrictions improve the significance of the model system, but as the model system only includes two commodities, these applied restrictions have strong impact on the results forcing the four price related parameters to have the same numerical size. By restricting $\gamma$, only the price elasticities are affected and differ quite a bit from the parameters of the unrestricted model system. Generally, the elasticities increase with the restrictions and travel packages become a substitute good for plane travel. By further adding restrictions on $\beta$ and $\alpha$, the income elasticities change such that plane tickets become a necessity.
6 CONCLUDING REMARKS

Using the Danish Expenditure survey from 1996 to 2007, this study analyses the relations within expenditures on transportation and leisure with special focus on long distance travelling represented by the expenditures on plane tickets and travel packages. The central question of interest was the elasticities of long distance travel to provide measurements for possible forecasting of the Danes total amount of travelling in relation to economic changes in society. This question does not seem to be commonly handled within the literature and the difficulties related to the variation in prices indicate why.

Both the aggregate demand system modelling with AIDS and the Tobit approach on the disaggregated commodities fall short on estimating sufficient price related parameters, and hence also price elasticities. However, the two applied methods provide realistic and robust income elasticities. Transportation and leisure as a joint commodity, as well as plane tickets and travel packages as separate commodities, are found to be luxury goods. The income elasticity of transportation and leisure is estimated to 1.2, whereas it is 1.1 and 1.4 respectively for plane tickets and travel packages. This implies that if the present growth in wealth continues, the population will generally travel more or e.g. increase the standard of travelling to higher luxury or more unique destinations.

A various number of estimations have been tested for the disaggregated commodities of plane tickets and travel packages, not all included in this paper. All the estimates show the same finding that travel packages are considered as a higher luxury than plane tickets and also a higher luxury than transportation and leisure in general. Even though the lack of variation in prices causes trouble in estimating sufficient price elasticities, the various results also indicate similar conditions that travel packages are far more sensitive to changes in prices than plane travel.

7 ACKNOWLEDGMENTS

The author thanks Ismir Mulalic and Jeppe Rich for helpful discussions on the methods applied and comments on the paper.

8 LITERATURE


Wang Yang, PhD Candidate,
Transport Research Centre (TRANSYT)
Universidad Politécnica de Madrid
34 91 3366708
wyang@caminos.upm.es

Improving the analysis of a toll ring scheme implementation by a Travel Demand Management model
Abstract:
The mobility survey concerning the Madrid Metropolitan Area (MMA) shows that from 1996 to 2004 the number of mechanized trips increased by 52%, whereas the population increased only by 14%. Moreover, even if there is a heavy investment made for public transport, its mode share still decreased in the last decade owing to the use of private car get more and more popular.

Since most countries developed extensive roadway networks during the Twentieth Century, the major transportation problems facing most communities are traffic and parking congestion, inadequate mobility for non-drivers associated with high numbers of private car trips; all problems which can be addressed by Transportation Demand Management (TDM).

An efficient strategy of TDM is the implementation of road pricing policies in the more congested road sections. One of the most popular “road pricing policies” is the toll-ring scheme used in Oslo and in Stockholm. In fact, a toll-ring scheme discourages orbital diversions, achieving higher traffic efficiency and environmental benefits.

This paper describes the application of a toll ring scheme in Madrid Metropolitan Area and realizes a comprehensive analysis for evaluating the toll-ring impacts in relation to the congestion relief and social equity. The simulation of the toll ring scheme implementation is realized by a travel demand model, by using the data from the household survey of Madrid in 2004.

Keywords: Madrid Metropolitan Area, toll-ring, efficiency, equity

1. INTRODUCTION

Following the previous study (Di Ciommo et al. 2010) of a toll ring scheme implementation on Madrid Metropolitan Area (MMA), this paper specifically focuses on the analysis of a toll ring pricing scheme in terms of traffic efficiency and social equity by using an original methodology.

In Madrid Metropolitan Area, 14.51 million trips are made in any weekday with an average speed 27 km per hour for cars in the year 2008 (Ministry of Transport of Spain, 2008). Like most of the metropolitan areas, Madrid is experiencing a heavy congestion issue, especially on the metropolitan rings of M-40 during the peak hours, which is accompanied by numerous
negative consequences, such as increased pollution, reduced road safety, social exclusion and generation of external costs, etc.

The most effective strategy to solve the congestion and its negative external consequences is to apply road pricing policy, which has been proposed for almost forty years as a practical means of reducing the externalities arising from road use (Ministry of Transport, 1964).

In general, the efficiency of a congestion pricing scheme is associated with the economic dimension. Most of researchers (Pigou, 1920; Button, 1998; Liu, 1999; Maruyama, 2007) have designed a welfare function to evaluate the impacts of a pricing scheme. However, this paper analyzes efficiency in traffic engineering sense by considering the traffic volume reduction and average speed increase (Viegas, 2001). Actually, we limit the analysis to the traffic efficiency (congestion relief) without evaluating the welfare function. We focus on an evaluation of the equity variation after implementing a toll ring pricing scheme, applying an inequality index like the Gini coefficient. In fact, the use of the welfare function to evaluate the equity effects could be problematic if we used a welfare function defined like the sum of the consumer surplus, like most of researches do (Kristoffersson, 2011). In this way, the welfare function is weighting more the wealthier people (Jara-Díaz, 2007).

Therefore, this study analyzes a toll ring pricing scheme charged on the orbital highway M-40 by using simulation tool VISUM in Madrid Metropolitan Area (MMA) to evaluate its impacts on mitigating the congestion (traffic efficiency) and the changes on social effects (equity) using the traffic data obtained from the household survey of Madrid in 2004 (Transport Authority of Madrid, 2006).

The choice of the M-40 ring is justified by two different reasons: first, it is the most congested orbital highway in MMA, and the first candidate for implementing a policy traffic measure to mitigate the congestion problems; second, as it can be seen as an external boundary, a toll of access to this ring can incentive a recentralization of the residential and economic activities inside of the ring. Like showed by May (2002), a toll-ring scheme intercepts more of the journeys which contribute to congestion and discourage orbital diversions, achieving higher efficiency and environmental benefits.

To support this analysis of the impacts of a toll ring scheme, the development of a model of driver route choice behavior is essential. The travelers in this model will then choose the path (route choice) that minimizes their generalized operation costs. A toll-ring scheme can be applied directly into this modeling framework. In particular, this study analyzes social equity by improved methodologies to evaluate more thoroughly the impacts of a toll ring scheme in Madrid Metropolitan Area.
CONGESTION PRICING POLICIES

Many municipal authorities have employed road pricing as a strategy of travel demand management with restriction of traffic inside of the city centers (Yamamoto, 2000). Virtually all of the studies to date have assumed that road pricing would be implemented through access points (cordon pricing). This can be done in a simple way, with a single cordon and charging in one direction only, as in Singapore (Rietveld, 2003). It is however possible to have a complex series of toll ring and cordons, with charges varying by direction and time of day; the Hong Kong and London road pricing implementation both provide examples of cordon pricing schemes (Richards et al., 1974).

However, such complex structures have been criticized because they are inflexible, since the fixed charging points cannot readily be relocated if the traffic conditions change; in consequence, these pricing schemes are unfair, in imposing the same charge both on longer and shorter trips; and that they can be disruptive, by producing more congestion on the boundary routes immediately outside the cordon road (Litman, 1996). These weaknesses all help to explain the marked sensitivity of performance to the design of the cordon pricing scheme; a recent London study found a threefold range of economic and environmental benefits across the range of cordon schemes tested (May et al., 2002).

In response to these criticisms, two other systems have been developed, with charges within a defined area based on time spent travelling (time-based pricing); and distance travelled (distance pricing). Different types of road pricing focus on different objectives, such as revenue generation or congestion management, or both. For example, road toll and HOT lanes help to raise revenues, cordon tolls and road space rationing help to reduce congestion in urban centers, and congestion pricing and vehicle use fees contribute to achieving a combination of reduced congestion and increased revenues (Blackledge, 2009).

The objectives more commonly pursued when implementing a pricing policy are: reducing congestion, improving the environment, and generating revenue (May, 2000). However, there are other important aspects to consider, such as urban sprawl or equity.

This last one is considered by several authors as one of the main obstacles for citizen acceptance of toll charge (Litman, 1996; Schlag, 1997; Link, 2007). Political acceptability depends very much on the distribution (and perception of the distribution) of gains and losses to a proposed change (Levinson, 2009). One of the ways to mitigate this effect is to use at least some of the revenues obtained, to improve the public transport system. Since that system is used disproportionately by those with lower incomes, this helps to transfer revenue from higher-income individuals to lower income ones, which means, according to the Dalton principle, an improvement in equity (Ramjerdi, 2006; Rietveld, 2003).

Transport Economics, Policy and Behaviour
This document is organized as follows: after the introduction and a discussion of congestion pricing scheme; Section 2 illustrates the methodology used to develop this research work and the simulation scenarios as well; all simulation procedures and contents are shown in Section 3; Section 4 explores the evaluation effects in terms of efficiency and equity; and finally, Section 5 presents the conclusions of the study by summarizing the findings of the study.
2. A METHODOLOGY TO IMPROVE THE ANALYSIS OF A TOLL RING SCHEME

The approach adopted is based on a travel demand management model defined by calibrating the simulator software, VISUM. Figure 1 shows the developed methodology of road pricing implementation analysis.

![Figure 1: Analysis Methodology Flow](image)

Before implementing a road pricing scheme in existing road network, it is necessary to study the current local situation, in this case, it means the circumstances of Madrid Metropolitan Area. In previous researches (Wang, 2011), the current situation study including socioeconomic information (population, average income, GDP, etc) and mobility indicators (such as traffic flow, travel speed, modal split and so on) has been studied, and in addition, a short discussion of the main causes and phenomena of local transportation problems has been presented as well. Following these researches, based on this information and the objectives of different pricing schemes; a toll ring scheme on M-40 is designed to solve the problems analyzed in the study area.

The second step is the simulation of road pricing scheme implementation. Via the designed pricing scheme, the corresponding simulation scenario is addressed. On the other hand, for executing a drive route choice behavior simulation with the software VISUM, several issues need to be prepared in advance. The simulation is performed by defining impedance function for each road section.

After that, the next step is the analysis and the evaluation of results. For evaluating efficiency, the traffic volume and average speed of a certain area (such as M-40 ring, other neighboring...
orbital highway) are obtained by comparison with a base scenario (without toll) to evaluate the impact on efficiency in terms of congestion relief.

Regarding the issue of equity evaluation, a method of calculating the GINI coefficient by including the individual income, travel cost (operation cost and toll cost) and time saving of each original zone is produced to compare the social equity of the pricing scheme.

**MADRID METROPOLITAN AREA STUDY**

Following the previous research on the value pricing experiment (Di Ciommo et al., 2011), the decision to analyze only the route choice changes of the car users, without considering the modal transfer or the time shifting, depends on the study of car dependency, road dependency and traffic dependency of the road network on MMA. Following some other research on the value pricing experiment (Lam, 2001), the decision to analyze only the route choice changes of the car users, without considering the modal transfer or the time shifting, is justified mainly by two different reasons:

1) The dependency on the car of the users of the orbital highway in MMA because of the radial form of the public transport networks and the urban sprawling of workplaces and residences that jointly increase the travel time ratio between public transport and car (see Figure 2 and Table 1).

2) The main travel motives identified (work, school or accompanying other people—mainly child to school-) make difficult to shift the travel hours, and thus lessen the congestion during the peak hour (8 am to 9 am). More than 90% of the travels are made for reasons related to working, studying or companying children to school, and normally the schedule of this kind of activity is fixed, so users with this travel motive cannot shift their depart time (see Table 2).

<table>
<thead>
<tr>
<th>Road Sections</th>
<th>Origin Municipalities with highest percentage of volume traffic of the selected section of the M40</th>
<th>Travel time ratio between modes (PT time/ Car time)</th>
</tr>
</thead>
<tbody>
<tr>
<td>North</td>
<td>Pozuelo de Alarcón 1.74 Rozas de Madrid (Las) 1.64 Majadahonda 1.70 Moncloa-Aravaca (district of Madrid) 1.61 Villanueva del Pardillo 1.91 Torrelodones 1.27</td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>Villa de Vallecas (District of Madrid) 1.91 Getafe 1.62 Villaverde (District of Madrid) 1.53</td>
<td></td>
</tr>
<tr>
<td>Road Sections</td>
<td>Motive of Travel</td>
<td>Percentages</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>North</td>
<td>Work</td>
<td>47.72%</td>
</tr>
<tr>
<td></td>
<td>Study</td>
<td>22.72%</td>
</tr>
<tr>
<td></td>
<td>Accompanying other people (mainly child to school)</td>
<td>19.49%</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>10.07%</td>
</tr>
<tr>
<td>South</td>
<td>Work</td>
<td>61.51%</td>
</tr>
<tr>
<td></td>
<td>Study</td>
<td>15.05%</td>
</tr>
<tr>
<td></td>
<td>Accompanying other people (mainly child to school)</td>
<td>15.58%</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>8.22%</td>
</tr>
</tbody>
</table>

Table 2. Proportion of travel motives in the selected road sections of the M40

Figure 2: Madrid Metropolitan Area with public transportation network

**INTERNALIZING EXTERNAL COSTS BY A CONGESTION PRICING SCENARIO IN THE MMA**

This paper mainly focuses on the congestion costs, which are the basis for the toll ring pricing scheme in a metropolitan area. And an average operation costs of each vehicle per kilometer estimated (0.1€/veh-km) by the META model for Madrid Metropolitan Area is used throughout this study (Di Ciommo et al, 2010).
Since this study is to simulate driver route choice behaviour during the peak hour (8 a.m. to 9 a.m.) in metropolitan area, the pricing scheme of each simulation scenario corresponding to the pricing scheme is based on:

- the pricing list of the existing toll highways in Madrid region, such as R-2, R-3
- the relative application pricing list in London, Oslo and Singapore.
- the previous study of pricing scheme implementation in MMA (Di Ciommo et al, 2010a, 2010b, 2011)

Table 3 lists the pricing scheme and its charging price corresponding to a given simulation scenario.

Table 3. Pricing list regarding to the pricing scheme

<table>
<thead>
<tr>
<th>Pricing Scheme</th>
<th>Price</th>
<th>Scenarios</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without toll</td>
<td>0</td>
<td>Scenario 0</td>
<td>Reference scenario, without charging</td>
</tr>
<tr>
<td>Toll Ring</td>
<td>2€/veh+0.1€/veh-km</td>
<td>Scenario 1</td>
<td>A toll ring scheme on the outer orbital highway of MMA</td>
</tr>
</tbody>
</table>

1. Scenario 0 (S0): Base case (without toll)

   In order to assess any type of policy or situation, a reference scenario or “base case” should be defined as a reference situation. In this study, it refers to this reference scenario (without charging or without toll) by the code S0.

2. Scenario 1 (S1): Toll ring based pricing in M-40

   A toll-ring scheme discourages orbital diversions, achieving higher efficiency and environmental benefits. It consists of two parts, a cordon toll on the access points to the ring delimited by M-40, while the kilometric toll is applied to all the links of M-40 to avoid congestion generating on the boundary of charging area. So a combined toll based on a cordon and distance scheme is simulated in the metropolitan highway M-40 during the peak hour (8-9 a.m.), which captures mainly the workplace and school trips.

The orbital highway M-40 is the main metropolitan ring surrounding Madrid with a total length of 63.3km, looping around the city at an average distance of 10.1 km to the city center. It is the only one of the several ring roads serving Madrid that functions as a full-fledged motorway for all its length (the inner ring, M-30, has a span about 2 km long at the northern arc that are not freeway-grade, having level crossings and traffic lights, and M-45 and M-50 are not complete rings). Some sections of the M-40 are amongst the most congested roads in Spain (23 km/h during peak hour). As a consequence, the scenario specifies tolling at the points of entrance and exit to and from the inner ring delimited by M-40 and at the links of the M-40 as well. The combined “access-and-at distance” based charges are estimated at around 2.0€+0.1€/vkm: 2.0€ represents the charge for entrance and leave to the inner ring and 0.1€/veh-km is the toll based on the distance traveled.
3. SCOPE OF SIMULATION

This part introduces the travel demand model, simulating the implementation of a toll ring scheme in MMA.

THE TRAVEL DEMAND MODEL

In this context, the tool used is a travel demand model that comprises the main road network of the region of Madrid, for private car travelers during the peak hour in a labor day (8-9 a.m.). The origin–destination matrix (O-D matrix) for the peak hour has been obtained from the Mobility Survey carried on by the Transport Authority in 2004 (CRTM, 2006). This survey has been conducted in Madrid region, interviewing around 35,000 families about their daily trip trips. The aggregate number of trips obtained from the database is based on the original municipality to destination municipality of Madrid Metropolitan Area, and then it is needed to be arranged based on the original zone to destination zone of VISUM. According to the function:

\[ t_{ij} = T_{ij} \cdot \left( \frac{PZ_i}{PM_j} \right) \]  

\( t_{ij} \): number of trips between each origin zone \( i \) and destination zone \( j \)  
\( T_{ij} \): number of trips between each origin municipality \( i \) and destination municipality \( j \)  
\( PZ_i \): number of population of zone \( i \)  
\( PM_j \): number of population of municipality \( i \)

The number of trips between origin and destiny zone has been finally obtained as an OD demand matrix to VISUM.

In addition, the OD matrix has been calibrated by using 394 traffic flow data from the Ministry of transport and infrastructures, the Regional Government of Madrid and the Council of Madrid. In Figure 3, the calibration shows a significant statistical convergence by the following adjustment function: \( y = 1.0471x - 120.98 \) and with a coefficient of regression (R2) of 0.90.
**SIMULATION IN PRACTICE**

The simulation is realized by the simulator software application VISUM, with the special feature that it is designed for multimodal analysis, as it integrates all relevant modes of transportation (i.e., car, car passenger, and trucks, buses, train, pedestrians and bicyclists trips) into one consistent network model.

Because this study intends to simulate the driver route choice behavior in the light of using a toll ring pricing scheme, travel impedance is defined as the key element to affect drivers’ decision. It means each private car driver only chooses the route with the minimum travel generalized cost. Then, the impedance function is defined by three elements: current travel time, toll fee and operation cost, as can be seen in function 4.3.

\[
I = C_T + (C_{TC} + C_{TO}) \cdot \lambda
\]  \hspace{1cm} (4.3)

- \(I\): value of impedance of each road section;
- \(C_T\): current travel time, given by the demand transport model;
- \(C_{TC}\): toll cost, refers to the road pricing scheme;
- \(C_{TO}\): operation cost, related with the travel distance as well, it is considered by 0.1 €/km;
- \(\lambda\): conversion factor of monetary unit (euros) into time unit (seconds), considering a value of time (VOT) of 9 €/h.

By defining the impedance function, all pricing schemes are converted into mathematic functions which are then identified and assimilated by the simulation software, and then the relevant results can be obtained according to the user’s needs. The simulation results and the method of evaluating the results are introduced in the following part.
4. **EFFECTS: EVALUATION AND COMPARISON**

**EFFICIENCY EVALUATION**

This part starts by evaluating traffic efficiency in terms of congestion relief in the case of toll ring pricing scheme.

*Congestion Relief*

Tables 4 and 5 list the results of the simulation obtained from the demand model VISUM in terms of traffic volume (vehicle-mileage) and current average speed for the area inside of M-40, all the charged road sections of the toll ring, a neighboring orbital ring M-30 and 13 radials highways.

Table 4. Hourly traffic volume (1,000 vehicle-mileages) variation

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Area Inside of M-40</th>
<th>M-40 Accesses</th>
<th>M-40 Ring</th>
<th>M-30 Ring</th>
<th>13 Radials</th>
</tr>
</thead>
<tbody>
<tr>
<td>S0 (Without toll)</td>
<td>2,782</td>
<td>262</td>
<td>774</td>
<td>330</td>
<td>2,174</td>
</tr>
<tr>
<td>S1 (Toll ring)</td>
<td>2,753</td>
<td>195</td>
<td>716</td>
<td>334</td>
<td>2,182</td>
</tr>
</tbody>
</table>

Table 5. Average Speed (km/h) variation

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Area Inside of M-40</th>
<th>M-40 Accesses</th>
<th>M-40 Ring</th>
<th>M-30 Ring</th>
<th>13 Radials</th>
</tr>
</thead>
<tbody>
<tr>
<td>S0 (Without toll)</td>
<td>29.2</td>
<td>48.9</td>
<td>50.7</td>
<td>53.0</td>
<td>64.1</td>
</tr>
<tr>
<td>S1 (Toll ring)</td>
<td>29.2</td>
<td>58.8</td>
<td>60.7</td>
<td>52.0</td>
<td>64.0</td>
</tr>
</tbody>
</table>

From the obtained results in Table 2-3 and Figure 4, several findings can be summarized as below:

1) Neither traffic volume nor average speed is modified in the area of inside of M-40. It is because this area is located by the metropolitan area of Madrid and has the higher density population, and the demand to enter or leave this area is constant in workdays for the commuting trips especially. Besides that, if the drivers intend to change their route avoiding paying the toll, they need to detour to a much longer path instead of the original one and a much higher generalized cost is thus generated. These two reasons lead to the traffic volume in the area inside of M-40 not having any significant difference comparing with the results without a toll.
2) When the results are evaluated in all the charged road sections (accesses and ring) of M-40, which the toll ring implemented, both of them present significant changes. For the traffic volume, there is a 26% decrease in the accesses of M-40 and 7% in the ring. For the average speed, two of the charging places yield similar effect, with a 17% increase in the access and 20% in the ring. These results prove the efficiency of a toll ring scheme in terms of congestion relief.

3) For the orbital ring M-30, which is the closest neighbor ring, there is no change shown in terms of traffic volume and average speed. This implies that the toll ring scheme implemented in M-40 may not cause traffic volume increase or average speed reduction.

4) The same result as found in the M-30 applies as well in 13 radials which are assumed to be the alternative routes for drivers. The traffic did not seem to be affected by the implementation of the toll ring scheme.

![Figure 4 - Traffic Flow Changing between S0 and S1](image)

**Equity Evaluation**

Equity refers to the distribution of resources and opportunities (Levinson, 2009). Traditional economic equity analysis supplies the notion of an income-distribution index, which is a measurement of income distribution within any given population. In transport, these measures have been occasionally adopted as part of the policy analysis. Vold (2005) adopted the Gini coefficient to evaluate the equity impact of different transport policy packages using Kolm’s measure. However, it is known that social equity measurement is very complicated depending on the studied area and target group (Lucas, 2006).
In this paper, we evaluate the equity by calculating the GINI coefficient in two different scales; macro scale is referred to the social equity in the whole transport network of Madrid Metropolitan Area, and the micro scale is considering the equity changing in two road sections of the charging area M-40 ring.

The Gini coefficient is a measure of the inequality of the income distribution, a value of 0 expressing total equality of the income between a population and a value of 1 meaning a maximal income inequality. It has found application in the study of inequalities in disciplines as diverse as economics. It is commonly used as a measure of inequality of income or wealth (Gini, 1936).

Two variables are necessary to calculate the Gini coefficient. One is the population of the studied area, another is the corresponding income. Since the individual income of drivers in Madrid Metropolitan Area is not available, the population and income based on the zone of VISUM are set to be the variables. And their value is obtained from the survey of Madrid region (Economic Research Institute, 2004). We develop two methodologies to analyze the social equity in macro and micro scale respectively.

**A methodology of equity evaluation at macro scale**

A methodology is introduced to explain how the income is affected by the time saving and toll spending in the whole transport network of MMA.

Since a pricing scheme is applied in the road network, the distribution of the residential income may produce an influence depending on the road user; the travelers could receive the changing of their income along with the variation of travel cost and travel time. Some of the travelers alter their path to avoid paying a toll; some of them remain on their route to save travel time. Two of these changing refer to the definition of impedance (generalized cost), so the changing of income is represented by the changing of impedance. Using the following function to express the relation between the income before and after implementing a pricing scheme:

\[
I_a = I_b + C_t + C_c = I_b + C_i
\]  

(5.1)

\(I_a\) : income after applying a road pricing scheme per year, in euros  
\(I_b\) : income before applying a road pricing scheme per year, in euros  
\(C_t\) : travel time changing, in seconds  
\(C_c\) : travel cost changing, in euros  
\(C_i\) : impedance changing, in seconds

Note: the impedance is calculated in terms of time for one peak hour of a labor day, it is necessary to transfer to monetary value for a year to the corresponding annual income.
By using the distribution of income ($I_a$) corresponding to the distribution of population of each zone, the Gini coefficient of each scenario is calculated and shown in the following table.

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>GINI Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>S0 (Without toll)</td>
<td>0.226</td>
</tr>
<tr>
<td>S1 (Toll ring)</td>
<td>0.155</td>
</tr>
</tbody>
</table>

The GINI coefficient of the toll ring scheme has been improved comparing with the reference value of 0.226. The achieved result of 0.155 shows the improvement of equity by implementing a toll ring pricing scheme in the whole transport network of MMA. It also proves the view of inequity can be mitigated through the basis of charging.

*A methodology of equity evaluation at micro scale*

At micro scale, we adopt a methodology to calculate the Gini coefficient in two different road sections located in the south and north parts of M-40, as shown in Figure 2. In this way, we can analyze the socio-economic effect of the metropolitan toll on users with different levels of income. Once categorized by their route choice after the introduction of the toll we examine the incidence of a congestion pricing policy on the income distribution of the users, comparing the Gini index in both the reference (without toll) and the toll-ring scenario.

After the toll-ring implementation, the Gini index shows a change (0.133-0.136) among the users of the northern road section. In this area of the network, the burden of the congestion toll has a greater effect on the lower-income users, as shown in Figure 5, while in the south, the burden of the congestion toll is more evenly distributed and the Gini coefficient is quite constant (0.131-0.132) with respect to the scenario “without toll”.

![Figure 5 - Gini coefficient level for each road section selected with and without toll](image-url)
A deeper analysis of the two different users population (north and south) shows a far location in the north for the low income population who has not the convenient alternative route. Even if the north population in general shows a higher level of income, the lower income population localized in the north pays proportionally more for the congestion toll policy because they live far from the alternative route and they are obliged to take the M-40. The impact of a congestion pricing scheme could be progressive on all travelers, but regressive on low income car owners (Banister, 1994)

Adopting the hypothesis of the inequalities of use (Coutard, 2002), we can conclude that in the south the implementation of a toll ring in the M40 affect users equally irrespective of income level. In contrast, the toll-ring on the M40 section in the North lowers social equity because of the lack of any alternative roads, and, since increased toll costs are much more burdensome to the lower-income users, social equity decreases.

By developing two methodologies in macro and micro scale, the equity analysis shows different results. In the macro equity study, the Gini index decreases after implementing a toll ring scheme, in higher scale, the social equity has thus been improved. However, we do not obtain the same result by the micro analysis by calculating the Gini index in two specific road sections. Particularly, the social inequity increases in the lower-income users who drive through the northern section of M-40.

Since there is not a general and authoritative approach to measure social equity, the conclusion quite depends on the methodology the researcher adopted, and depends on the studied area as well. In this paper, we just show the changes in equity through two different methods by a simulation result. Certainly, in real life, social equity has to be measured regarding to the stakeholders’ specific objectives, such as car dependency reduction or social exclusion mitigation, etc.
5. CONCLUSIONS

This section summarizes the key findings of the present research work, whose main objective is to assess the effects of a toll ring scheme on efficiency and equity based on the simulation in road transport network of Madrid Metropolitan Area. Decision makers may integrate in their policies different measures that can increase pricing scheme effectiveness, like showed by the results of the application of the methodology.

The key findings obtained from the application of the effects evaluation to the case study of Madrid Metropolitan Area can be summarized as follows:

1. In terms of congestion relief, the traffic efficiency analysis shows that the toll ring scheme charged on the orbital ring M-40 can generate a distinct traffic volume reduction and average speed increase. But it does not relieve the congestion inside of M-40, mainly because of its big size and single travel mode. At the same time, it can be seen that this toll ring scheme does not cause a transfer of the congestion to the neighboring highways (M-30 or 13 Radials), leading to additional congested highways.

2. As important as efficiency, the equity analysis, which is mainly considered by calculating the GINI coefficient both at macro and micro scale, generates two different results. Equity is improved after applying a toll ring pricing scheme in M-40 in the macro analysis. But it shows a different result in the micro study which evaluates the social equity in two road sections of M-40. Therefore, we cannot conclude that a toll ring scheme can improve the equity. It can only be measured by the stakeholder regarding to the specific objectives in real life.

Overall, the simulation procedures applied in this work present the general impacts with respect to efficiency and social equity. A toll ring pricing scheme simulated in the road network of Madrid Metropolitan Area might produce an effective result to relieve the congestion and promote social equity. However, the performance in terms of efficiency and equity depends on the level of pricing and the charging area. The results obtained by the simulation procedures can be readily used by transport authorities and managers in order to bridge the gap between state of the art and state of the practice.
REFERENCES


Blackledge, D et al. (2009). CURACAO: Coordination of urban road user charging organizational issues, Leeds, 2009


Ministry of Transport of Spain, (2007), The Metropolitan Mobility Observatory.


Author: Jørgen Aarhaug

Institute of Transport Economics
jaa@toi.no

Competitive tendering in an entry regulated market – an accident waiting to happen?
Abstract

Competitive tendering has increasingly been used in procurement of public transport, with well documented effects in lower costs and more clear distribution of responsibility. The effects from competitive tendering on rural taxi service are much less documented.

From 2003 onwards competitive tendering has been introduced in the public purchase of taxi services. The intended effects include lower prices for the purchasers and a more efficient use of public funds as the responsibility for the transport of patients has been moved from the counties to the newly formed health corporations. This paper will discuss the anticipated and observed effects from this transfer of the tendering model into a new sector. In particular this paper will discuss some of the hypotheses put forward by involved actors, such as competitive tendering secures lower prices and better service, competitive tendering results in lower prices and reduced service, competitive tendering results in an unsustainable industry.

This paper is based upon data from recent research on the Norwegian taxi industry. This includes detailed datasets on particular cases and a more general data for the entire country, providing an overview of the situation.

Main findings presented in this paper are that at present there is little real competition in the Norwegian rural taxi markets, the process of introducing competitive tendering has not resulted in many new entrants to the taxi markets. However there are some cases where there has been real competition. In these cases this has resulted in different outcomes. One explanation for the very mixed results from introducing competitive tendering to the rural taxi market seem to be that there is a conflict in the regulatory regime between the entry regulation of the taxi industry and the need for more entrants in order to have sufficient competition for a regime based upon competitive tendering for the procurement of rural transport services.
Introduction:

Competitive tendering has increasingly been used in procurement of public transport in particular local bus services, with well documented effects in lower costs and more clear distribution of responsibility. The effects from competitive tendering on rural taxi service have received far less academic attention.

This paper provides an overview of the Norwegian taxi industry, and an analysis of the effects from introducing competitive tendering on public contracts has had on this industry, three possible explanations are presented and discussed in light of theoretical expectations and empirical observations. Particular emphasis will be placed on the explaining power of the different hypothesis in rural areas. Rural areas has several aspects which make them an interesting case for study, the markets are smaller, there are fewer potential entrants and there will be some form of interdependency between the taxi industry and their customers. All of these are aspects that will to some extent affect the outcome of the tenders. As rural areas are perceived to be more vulnerable, and there is a political will to maintain service levels, rural areas also have received more systematic study than taxi services in urban areas. This paper is to a large extent based upon material from Osland et al. (2010).

As background information this paper is first providing an overview over the Norwegian taxi industry, including a short presentation of the regulatory regime. This is followed by the changes in the regulatory regime from year 2000 to present. Further on expectations from theory and different explanations to the processes at work as a result in the change in the regulatory system is presented. The empirical relevance of these is discussed by using the actions taken by the different involved parties as indicators. This paper is concluded by pointing at the merits of the different explanations and possible development trends.

The Norwegian Taxi industry

Norway’s 4, 9 million inhabitants are served by approximately 8898 taxis (Longva et al. 2010) employing about 16000 persons (Egeland et al. 2009). As these numbers indicate, the Norwegian taxi industry is very fragmented and dominated by owner-manager businesses, with between one and three employees, including the license holder. There is a requirement for one license per vehicle and the rule of thumb is that there is one license per license holder. Although there are exemptions made from this rule there are very few cases where the license holder holds more than two licenses. Taxi density, measured as number of vehicles per 1000 inhabitants is highest in urban areas (Osland et al. 2010).
As the taxi market is divided between many different segments, revenue is also divided between different sources. In urban areas private customers dominate, this includes taxi rank, telephone booking and private contracts. However, even in urban areas public contracts account for between a third and a half of total revenue; this includes contracts on school transport, transport of the disabled and transport of patients. From the limited data available it seems that the transport of school children and patients are the two most important. In rural areas public contracts dominate and the transport of patients is clearly the most important contract in monetary value (Osland et al. 2010).

The majority of taxi licenses in Norway are affiliated to a dispatcher (88%). The dispatching companies have a mix of ownership structures, the two dominant forms are joint venture between license holders and limited companies specializing in dispatching services (Longva et al. 2010). However there are examples of integrated companies, which provide both dispatch and operation.

*Regulatory regime*

Responsibility for taxi regulations is divided between three authorities. Overall responsibility lies with the ministry of transport and communications (MTC). MTC also stipulate quality and safety requirements for the license holders and regulate market access for operations crossing county lines. The Norwegian Competition Authority (NCA) enforces the Norwegian Competition Act and sets the maximum fare. Market access regulation and licensing, both for taxi licenses and dispatchers is conducted by the County Governments (CG) in each of Norway’s 19 counties, this is done under the provisions of the Professional Transport Act. The licenses are only valid in the county where they are issued.

The Norwegian taxi regulations of today can be traced back to the 1940s. The pivot of this regime is the entry regulations, in the form of a needs test. This has traditionally been seen as a way of securing adequate service for the public while maintaining good working conditions and secure income for the taxi owners. The entry regulation is administrated by the county government, in accordance with the Professional Transport Act (Aarhaug and Osland, 2010a).
Figure 1. Entry regulation balancing supply and demand side interests (Aarhaug and Osland, 2010a)

Figure 1 illustrates the important role played by entry regulation in the Norwegian taxi regulations. Entry regulation is the tool used by the CGs in order to balance the interest of the taxi industry against the interest of the general public. On one hand, passengers are guaranteed a certain level of service at predictable and reasonable prices through the license requirements. On the other hand, the entry regulation element is supposed to secure a reasonable income and working conditions by preventing what the industry has labelled “cream skimming”, that is taxis only available on high demand periods.

There are two different types of regimes regulating the Norwegian taxi industry, one for the six largest cities including surrounding areas (urban areas) and one for the rest of the country (rural areas). The main difference between these is that there is no maximum fare in the cities while the rest of the country has a maximum fare on the single trip markets. In both areas entry regulation is practiced. In the Professional Transportation Act entry regulation is seen as balancing the interests of the taxi industry (income security and working conditions) against the public interests (of 24-hour service and modest prices).

Qualitative requirements are set by the MTC, but the MTC has little part in enforcing these regulations. That is delegated to the CGs, who usually does this
together with the quantitative regulation by only giving licenses to qualified persons. However this part of the regulation needs revising as acknowledged by the minister of transport and communication (Kleppa, 2011).

*Changes in regulatory regime*

The split between areas with and areas without a maximum fare in Norway dates back to the year 2000. Before this the maximum fare was applied also in the largest cities. After the six largest cities with surrounding areas has been exempted from the maximum fare. In these areas the dispatchers are now able to set their own fares in the market above the maximum fare. In these areas there is as a rule also more than one dispatcher in each area. As the most important economic unit is the owner-operator these are allowed to give discounts from the price marketed by the dispatcher. This is however rarely done. In the rest of Norway, where the maximum fare applies the maximum fare is still only in use for the single trip segments, i.e. the hailing, rank and telephone market segments. This means that income from fare regulated driving only accounts for a small percentage of total revenue, even while only looking at the areas where it applies. Nevertheless the maximum fare is used as a reference in many contracts, giving the regulated maximum price far more effect than the market segments to which it applies indicate (Aarhaug og Osland 2010a).

Today, both in urban and rural areas prices in contracts are usually a result of competitive tendering in some cases followed by negotiations. This is a change compared to ten years ago, when negotiation dominated and competitive tendering was hardly mentioned. This change has occurred on most contracts for the public sector and can be seen as a reform in its own right, but for most of the contracts this has been a result of local, not national decisions (Osland et al. 2010).

In Norway there was a major reform of the public sector, which was implemented in 2004. Motivating this reform was the possibility to utilize the limited resources available by restructuring parts of the public sector. The sector which received most attention was the healthcare sector. Responsibility for the hospitals was transferred from the county governments to the newly created healthcare corporations (Helseforetak), which in turn were grouped into regional healthcare corporations. As a part of this reform responsibility for the purchase of healthcare related transport services, (the non-emergency transport of patients to treatment and between hospitals), was transferred from the counties to the newly created health corporations. The purpose was that the health corporations could
internalize transport costs in their treatment decisions, and thus achieve more efficient utilization of the resources available.

For the case of the transport of patients this was to be done by utilizing competitive tendering. It is a bit unclear what the expectations for this were. Osland et al. (2010) observe that a main focus from the health corporations has been to reduce the total cost of transport, not cost per unit of transport. Statistics available do indicate that control over total costs, at least partially has been achieved, but it is not possible to derive what has actually been produced. Therefore it is very difficult to establish how this potential saving has been made. From the NCA there seemed to be an expectation of lower prices. This can be observed on some contracts, but on most contracts the opposite has been observed. Very few tendered contracts have received more than one bid, and in particular in rural areas there have been very few non-taxi entrants on these tenders.

*Expectations from theory*

For public transport there are very clear expectations, both from auction theory (McAfee and McMillan, 1987), and from the empirical research done on competitive tendering with focus on the bus sector. The expectation from auction theory is that the price will be equal to the willingness to pay for the person/actor with the second highest willingness to pay for the good in question. This prediction is the same in most common forms for auctions (first price sealed bid, second price sealed bid, English and Dutch), but the expected variation differs. Competitive tendering can be seen analyzed as a reversed auction. In competitive tendering the good is the contract and the price the subsidy. Expectations are for lower subsidies as the number of entrants to the competition increases. The result is lower prices and as a consequence, more efficient use of public resources as a result of competitive tendering. These expectations are well documented for public procurement of bus services in Scandinavia.

Benefits from introducing competitive tendering in public transport are well documented, by amongst others Bekken et al. (2006) using Norwegian data and Alexandersson and Pyddoke (2003), using Swedish data. The main benefit has been in the form of lower prices. Together with these well documented benefits it has also been a restructuring of the bus industry into fewer and larger companies competing on a national and international, rather than local basis, Mathisen and Solvoll (2008) and Aarhaug (2009). Over time the development is that competition has been working, with several competitors entering each competition and prices remaining low also in the second and later rounds of tendering.
(Alexandersson and Pyddoke, 2003) etc. Similar studies, of the effect of competitive tendering, on the taxi industry have not been done in Norway.

There are several reasons for why this model based upon competitive tendering can be less suited for purchasing taxi services compared to bus services. First the taxi market is regulated and organized in a way which limits the usefulness of this model. There are several problematic issues related to entry regulation and industry structure. A second is the size of the relevant markets and third the geographical context including the element of interdependency.

First, looking at the way the market is regulated there are several points in the regulation which can be at odds with a functioning market structured around competitive tendering. Primarily the linkage between entry regulation and needs tests, which in can be interpreted as an entry barrier. When the supply is limited the normal requirement in a competitive market, open entry fails. As there are several markets where there are entry barriers, but still functioning markets this does not in itself need to result in the competitive tendering model being rejected. But, entry regulation can in cases severely limit the number of competitors. As there are very clear economics of scale in some of the relevant market segments, due to high fixed cost. There can be argued for an “optimal solution” to be a monopoly or a duopoly, which will make the competitive tendering model, for the purchase of services less useful.

Second, looking at the size of the markets in question there are issues. Usually the tenders are divided into one or a couple of packages per municipality. The motivation for this is that in the absence of larger actors dividing the market into smaller packages allow smaller and also non-taxi actors to participate. However in the Norwegian competition act, there is an exemption made for taxi dispatching. Taxis are allowed to cooperate and organize dispatchers in order to improve efficiency (NCA, 2008). In other words there is probably a minimum size below which the gains from competition may be lower than the loss from lack of scale. What this size is probably a function of the technical requirements for the dispatchers and the fixed costs associated with providing such services. In practical terms, who pays for the planning part of the dispatching?

Finally the geographical context is very important. Rural Norway includes areas which has very low population densities and long distances to hospitals. Low population densities and long distances are linked with more limited services, and higher costs per trip. This will again result in a limited market for private customers as the prices for a given trip by taxi will be much higher than by private car. And also a limited market for business contracts, as the number of firms is limited. If the price is set to the expected costs of a taxi located more than 50 km away. This will allow for a price significantly above the regulated maximum.
price, particularly if both parties have an understanding that the price cannot reach a choke point. If competitive tendering is used, in such circumstances the expectation from auctioning theory is that the prices will be very high. However it is also reasonable to assume that it is known to the purchaser that the local taxi industry is dependent upon this contract for its existence and therefore, the power relations in a negotiation (not competitive tendering) will be unclear, the purchasers will have significant leverage.

In summary there are several reasons for why a purchasing model based upon competitive tendering that has proven to be very useful in other sectors is not directly transferable to the rural taxi sector, these reasons include regulatory issues, issues related to industry structure and issues related to the size and geographical characteristics of markets where the rural taxis operate.

**Hypothesis and theory**

From economic theory, and empirical experiences from other and relatively similar sectors, the general expectation should be that competitive tendering result in lower prices and better service, provided that sufficient competition exists and the tendering process is well conducted. In other words given that the conditions for competitive tendering applies the hypothesis stating that expectations should be lower prices and better service should be relevant. The question that remains to be answered is whether or not these conditions are in place.

The hypotheses put forward by some actors, that competitive tendering results in lower prices and lower services rests on the assumption that the tendering processes favours operators who achieve low costs through reducing the quality of the service. This again rests on the assumption that the contracts offered do not pay enough attention to quality requirements. From a theoretical point of view this can very well be the case.

The hypothesis that competitive tendering is resulting in an unsustainable industry is also resting on bad practises. And may very well be a theoretical possibility. The argument is that competitive tendering results in lower profitability on the contract markets. This in turn result in actors leaving the industry, as there are few available sources for extra income and the requirement for 24 hour service remains. Still for this to be the case there must be real competition on the tendered contracts, and if that is the case. One would expect the second round of competition to have fewer entrants and higher prices.
Hypothesis and empirics

There are very few places where competitive tendering has resulted in lower prices in rural areas. Also in urban areas there have been problems with attracting enough attention to have real competition; there are urban areas without real competition. Looking at only rural areas there is some examples where there has been real competition and in these areas there are reported lower prices and equal or better service. This is however exemptions not the rule (Osland et al 2010).

Osland et al. (2010) and Aarhaug and Skollerud (2011) find little support for the hypothesis of lower prices and reduced service. The data suggests that, provided the requirements in the tenders are set reasonably most actors, also non taxi, are able to provide the required level of service. And as stated by several informants there is no support for the claim made by loosing contenders that the winner has won by allowing services to passengers to drop.

Experience presented in Osland et al. (2010) suggests that there may be a real possibility that the industry is regulated in an unsustainable way. Looking at the cases in rural areas where there has been real competition, there has also been a reduction of the number of active taxi licenses. Osland et al. (2010) points at conflicts in the regulatory regime as an explanation for this development. In particular the combination of the licensing requirement and the new regime for public procurement is troublesome. The licensing regulations are based upon a system of regulated monopolies, while the procurement regime is based upon competition.

In the field solutions

As the different actors involved, in particular the different procurement agencies have encountered real or perceived problems where the map does not fit, they have found local solutions. In particular the question of limited competition has received a lot of attention.

One solution to the problem, perceived or real, of high fixed cost and limited number of entrants, has been to transfer the responsibility for trip organizing from the taxi dispatchers to a special unit at the health corporation. Similar steps have also been taken by several county governments. Organization is very labour and technology intensive and accounts for much of the fixed cost of taxi operations. The transfer of responsibility for these tasks was in many cases predated competitive tendering. By organizing the health related trips at health corporation level, rather than county or dispatcher, the health corporation can obtain full insight into the processes and at the same time this transfer reduces the requirement to and there by remove a significant fixed cost from the taxi
dispatcher. However, the taxi dispatcher is still required to have a 24 hour dispatching service, and it can be argued that this is in reality increasing total costs by duplicating services rather than reducing costs by increasing the number of potential competitors.

One documented result from this transfer of responsibility is that holders of licenses for occasional passenger transport, a license which only is subject to quality requirements not quantitative control, are gaining market shares. Before this reorganization there were very few, if any, non taxi operators on public contracts for door to door transport. Operators working on licenses for occasional passenger transport are however required to use vehicles registered for nine persons or more, including the driver. This limits the direct competition between these vehicles and small taxis. But the competition between these vehicles and taxis registered under the same requirements is very direct. There is also indirect competition, especially on longer trips, with the smaller vehicles, as a larger vehicle is able to pick up passengers en route and they have capacity to transport more passengers per trip than the smaller vehicles. In other words; elements of the problem associated with entry regulation, is at least on paper, abated by transferring responsibility for trip planning to the health corporations. This process has partly reduced the requirements for the operators allowing also license holders operating without entry regulation to compete for the contracts.

Another solution proposed to the problem of having few competitors has been to reduce and simplify the contract requirements. Reducing the barrier to entry the tendering documents them self has represented. This has only been done by some health corporations and on a limited scale. The reasoning behind this solution is that if the taxi owners do not need to cooperate in order to submit a bid, they choose to submit a bid that breaks with the dispatchers bid and the number of entrants is increased. At present there is only limited documentation on the results from this approach and this is mostly inconclusive.

Discussion

There is mixed experience from reducing the size of the tendering contracts. Most health corporations now operate with several packages for each municipality, usually distinguished by size of vehicle and if the trip is wholly within the municipality or not. In several cases more than one entrant has signed a contract. However the rule of thumb is still, as stated in Osland et al. (2010), that most packages still receive only one bid. Also in the cases where there have been more than one entrant, the smaller, and in many cases cheaper entrant, has not had sufficient capacity to fill the entire contract and the purchaser has signed contracts
with both entrants in order to fill their capacity requirements. As both the size of the package, and the size and capacity of the entrants can be assumed to be known, one will expect that the effects from reducing the minimal size in the tenders only will have limited effects.

Data presented in Osland et al. (2010) show that the income per license on average is significantly lower in rural, compared with urban areas and also show that the proportion of turnover from the non-public market segments is much smaller. As the municipalities and counties also in rural areas have an obligation for the transport of school children, and the health corporations for the transport of patients, there is still a demand for a taxi service. And there is interdependency between the taxi industry and public sector. The public sector is required by law to provide certain services, and the taxis do not have other major sources of income and is thus dependent upon the public contracts.

An overall view of the empirics show that some of the theoretical problems associated with competitive tendering in the rural taxi market have been addressed. Still it is, at best, unclear if these fixes have succeeded in creating a market in which sufficient conditions for a working regime based upon competitive tendering is in place. The market has been opened up for non-taxi entrants, but it is unclear where, in rural areas, there is a market large enough for both a taxi and a non-taxi operator. So far there has not been a general reduction in service levels in rural areas. There have been very few changes in the industry since the reform (Osland et al. 2010). This is a paradox as one from theory would expect that introducing competition on such markets either should result in the number of actors going down, as fewer than all will get a contract, or prices going up dramatically.

Neither has happened, however in certain cases prices has been increasing. A possible answer to the question why is that there has been no real competition, but at the same time, there is interdependency between the actors. And as a consequence of only receiving a single bid, several purchasers have entered negotiations. In other words, it can be claimed that service levels have been maintained as a consequence of the reform not meeting the targeted level of competition. If there is to be competition also on the rural taxi contracts in the future, the public will have to pay a direct subsidy for 24 hour taxi service, alternatively one can accept to not maintain this service or accept having very limited competition for the public tenders.
Conclusion

There is a contradiction in the taxi regulations in Norway. On the one hand there is a market based system with competitive tendering on public contracts, on the other entry regulation is retained together with regulated fares and obligations to perform a 24 hour service. This only has limited effects in the larger cities, where the fare is deregulated in all market segments and there in general is more than one actor who can provide the capacity needed on each contract. In other words there are reasons to believe that a regime based upon a form of competitive tendering will work well.

In rural areas this conflict is more problematic. There are several reasons for this. One is that the tendered contracts are a larger part; often well over half, of total revenue for the taxi industry. A second is that the size of the taxi dispatchers often is much smaller than estimated smallest efficient size. A further reason is that the number of actors is limited. However so far the problems associated with competitive tendering in areas with limited supply has only to a limited extent resulted in a reduction of the number of taxis. The main explanation for this seems to be that so far there has been little competition in rural areas. This is perceived to be a problem, but if this problem is solved and the result is real competition, it is very hard to argue for maintaining the 24 hour requirement and at the same time expect the number of vehicles available to remain or increase and the prices to fall. An efficient price on the competitive tendered contract may not be a price the public is willing to pay. And it is unclear if the total cost of the services now offered by the rural taxi industry will increase or decrease.

References


Bekken, J-T et al. (2006): ”Norwegian experiences with tendered bus services”, European Transport, Year XI, Number 33, August 2006.


NCA, 2008: “Unntak fra forbudet om konkurransebegrensende samarbeid”, (exceptions from the ban on competition reducing cooperation). http://www.konkurransetilsynet.no/Global/Faktaark/%c2%A710_UNNTAK.PDF

NCA, 2009: “Pasienttransport og konkurranse i drosjenæringen”, (Norwegian, ”transport of patients and competition in the taxi industry”) letter from the Norwegian Competition Authority 2009/245.


Determinants of Capacity Utilization in Road Freight Transportation

Megersa Abate*

DTU Transport, Technical University of Denmark

May 19, 2011

Abstract

In this paper we study two aspects of capacity utilization, namely the extent of empty running and the load factor. We show that they are explained as a function of truck, haul and carrier characteristics. We use a unique dataset from the Danish heavy vehicle trip diary that has detailed information about operation at a trip level. Our econometric model corrects for potential sample selection bias by jointly estimating the load factor and empty/loaded movement decisions of carriers. The results indicate that trip distance and being a for-hire carrier have a significant positive effect on capacity utilization. Interestingly, the effect of the carrying capacity of a truck on utilization is negative and non-linear.

*I am very grateful to Gerard de Jong and Tony Fowkes, whose guidance and suggestions have greatly improved the paper during my visit at Institute for Transport Studies (ITS), University of Leeds. I also thank Ole Kveiborg and Bertel Schjerning for their helpful comments on earlier drafts. All errors are my own. E-mail maa@transport.dtu.dk.
1 Introduction

Road freight transportation provides a vital service that ensures movement of goods within and between most economic sectors. The level of capacity utilization of trucks used for freight transport, therefore, shows how well economic resources are used from the perspective of transport operators and of sectors reliant on their services. Knowing how efficiently trucks are used is also central to the debate on sustainable transport since trucking operations are a significant source of harmful emissions to the environment. A recent performance report for 13 European countries reveals that about 30 percent of all trips made by trucks are empty, while for loaded trips the average load factor is about 50 percent over the period 1990-2008 (European Environmental Agency, 2010). Judged by these performance figures, it appears that more can be achieved by improving capacity utilization within the trucking industry. Improving utilization becomes all the more important if there are structural problems such as infrastructural bottlenecks for modal substitution (Rich et al., 2011).

Previous studies identified information technology capability of trucks (Hubbard, 2000, 2003; Barla et al., 2010), traffic composition (Boyer and Burks, 2010) and restrictive regulatory regimes (Wilson and Beliock, 1994) as the determinants of capacity utilization. The two main indicators of capacity utilization analyzed in the studies are the extent of empty runs and the load factor. Though insightful explanations for their determinants are given, the indicators are studied independently, and the explanations for why they differ between carriers are based on a rather aggregate data that hides how the actual freight movement took place. We often see loaded and empty trucks moving between the same origin and destination, revealing that carriers differ in capacity utilization level or face diverse market conditions. Little is known about what explains such differences, and as such, further studies are required to understand the resource allocation process in trucking.

The present study employs an econometric framework that simultaneously estimates a market access decision (empty or loaded movement decision) and the load factor (the level of capacity utilization during a loaded trip). The model is based on Heckman (1979), and it corrects for potential sample selection bias by explaining the underlying determinants of the level of the load factor and whether trucks move loaded or not in a joint estimation framework. Our detailed trip level dataset also provides us with an opportunity to apply a sensible econometric approach by controlling for key explanatory variables without aggregation biases. The data is from the Danish heavy trucks trip diary, and it has detailed information about operations at a trip level for each quarter of 2006 and 2007. We show that trip distance, carrying capacity, carrier type and other haul, truck and carrier characteristics are the underlying determinants of capacity utilization.

The rest of the paper is organized as follows: section 2 gives background to our analysis by reviewing the literature and section 3 presents the data and descriptive statistics. Discussion of the econometrics framework and results are given in sections 4 and 5, respectively. Finally, conclusions and recommendations are given in section 6.
2 Capacity utilization in road freight transportation

2.1 Definition and measurement

Capacity utilization is defined along physical dimensions and it mainly takes into account the physical (engineering) aspect of transportation service rather than a direct focus on firms. From the outset, it is important to note that the main focus of capacity utilization is physical productivity rather than economic productivity since price and cost are not directly considered (Nelson, 1987). The physical aspect of production in road freight transport involves a simple process of moving a load from one point to another by a truck and a driver.\(^1\) The production technology is a simple Leontief type technology where a single labor (a driver) and capital (a truck) combination is always required to produce an output (a movement of load). As pointed out by Boyer and Burks (2009), the scope for getting more output (e.g. tonne-kilometers) per truck per year using the classic manufacturing method - i.e. substituting capital for labor through automation - appears to be very limited.

It is interesting to note a unique feature of transporting freight compared to passengers and its implication to utilization. Unlike passengers, freight rarely returns to its origin.\(^2\) It usually moves from production areas to consumption areas via distribution or consolidation centers. On the other hand, a trucking service is by nature a joint-product; a truck movement from origin ‘A’ to destination ‘B’ (a front-haul direction) is usually followed by a reverse movement from ‘B’ to ‘A’ (a backhaul direction) or to the terminal of the truck carrier if it is different from ‘A’. Optimal capacity utilization depends on whether the truck is moving empty or loaded as well as the percentage of its carrying capacity that is filled with a load during its movement between these locations. The unidirectional movement of freight, however, implies less-than-optimal capacity utilization especially during the backhaul (return) trips.\(^3\) The problem is referred to as "the backhaul problem" in the literature (Ferguson, 1972; Felton, 1981; Wilson, 1987; Demirel et al., 2010).

Hubbard (2003) discusses two measures of capacity utilization. The first concept relates to the share of "loaded kilometers", defined as the numbers of kilometers trucks are driven with a load during the periods trucks are in operation away from their base. Seen this way, the performance of a truck can be evaluated by the share of "loaded kilometers" relative the total kilometers the truck is driven. The second concept considers the number of times trucks are in use in a period. For instance, trucks that are driven more weeks in a year are considered to have a higher level of utilization than those used for a fewer number of weeks. The extent of empty running is another important indicator of capacity utilization. A truck moving in a round trip between two locations may choose to carry a load on a front-haul direction and run empty during

\(^1\)The overall freight moving activity, however, involves a complex interaction between carriers, shippers and logistic service providers (see Holguin-Veras (2009) for an interesting discussion).
\(^2\)Trucks are also engaged in reverse logistics, but it constitutes a very small percentage of the total freight movement.
\(^3\)A study by Barla et al. (2010) finds both theoretical and empirical results which indicate that less-than-optimal capacity utilization can also occur in front-haul directions as carriers, helped by a truck’s information technology capability, move unloaded trucks in a front-haul direction in anticipation of getting a backload.
a backhaul movement and vice-versa. For a truck moving in a more complex pattern, empty runs occur at several stages of its journey. Empty running reflects sub-optimal capacity utilization if it arises due to a matching problem between demand and supply. In some situations, however, empty runs are inevitable. For example, as a result of geographical imbalances in freight movement, carriers may tend to move their trucks empty during a backhaul trip from a net freight importing region. McKinnon and Ge (2006) also identify the lack of transparency in the road freight market, short haul lengths, scheduling constraints and the incompatibility of vehicles and loads as possible causes of empty running.

### 2.2 Previous empirical studies

Previous empirical studies mainly focus on the cost structure of carriers to explain differences in utilization performance (see Beilock and Kilmer 1986 and Wilson and Dooly 1993 and Wilson 1994). Their basic explanation is that if there is a systematic access cost differential between carriers, we will see difference in utilization level between similar trucks even if they are used for the same haul and between the same origins and destinations. There are two implied assumptions in these studies, carriers incur more or less equal cost of operating empty trucks and they face equal freight rates. With regard to costs associated with accessing a particular freight market, however, some carriers may have cost advantages. Two main sources of market access cost advantages and their implication to trucking efficiency are discussed in the literature.

The first source of cost advantages arises from government intervention in the trucking industry. The effect of market regulation on carriers’ decision to access different freight markets is discussed by Wilson and Beilock (1994), Wilson and Dooly (1993) and Beilock and Kilmer (1986). Even if trucking activities are now free from government intervention in most countries, these studies identify an important source of variation in access cost. All argue, from different angles, that having a special license to haul a regulated commodity improves utilization (less empty runs) by lowering access cost; whereas, lack of the license forces carriers to forego loading opportunities. One important limitation of these studies is that utilization is measured solely based on whether a truck is loaded or empty without considering the load factor.

The second source of cost advantage between carriers is information technology (IT) capability of trucks. Controlling for carrier, truck and haul characteristics, Barla et al. (2010) and Hubbard (2000, 2003) show that IT capability results in higher capacity utilization by lowering search cost and improving the process of matching a vehicle capacity with available load. The specific aspects of capacity utilization considered in these studies are the load factor and the number of loaded kilometers, respectively. Both studies use data from the late nineties when IT adoption was relatively small in North America. Thus IT capability may no longer explain utilization differences since the technology might have diffused well by now. It is, therefore, likely that other attributes of trucks (size, fuel efficiency etc) or structural changes in the industry are playing important roles in determining how well a particular truck is used. For instance, Boyer and Burks (2009) argue that trucking in the US has increased its proportion of traffic that is relatively cheap to handle along with the share of long-distance hauls. Their finding implies that changes in traffic composition are the main
reason for the apparent productivity increase in the US trucking industry, ruling out “real” productivity gains resulting from systematic cost differential between carriers.

In another strand of literature that deals with shipment size and mode choice, the issue of capacity utilization is raised indirectly and the reason why it varies between carriers is given little attention. For instance, the interaction between shippers and carriers and its implication to mode choice has been extensively studied (see McFadden et al., 1985; Abdelwahab and Sargious, 1992; Abdelwahab, 1998; de Jong and Johnson, 2009; Holguín-Veras et al., 2009). The main finding in the studies is that the interaction leads to simultaneous choice of mode and shipment size to ensure that freight is carried by an efficient mode of transportation. In each chosen mode, however, carriers are simply assumed to allocate their fleet efficiently across hauls. Our analysis is an interesting extension to these studies’ joint empirical framework to understand efficiency at operational level.

3 Data and descriptive statistics

The data used for our study is from the Danish heavy trucks (with a maximum legal carrying capacity of 6 tonnes and above) trip diary that has detailed information about operations at a trip level. This level of disaggregation is rare and provides a significant advantage in studying capacity utilization. The diary is filled in by drivers for one week of operation (see DST, 2011 for details). We have such a dataset for each quarter of the years 2006 and 2007. The data is best described as repeated cross-section since the travel diary is filled out for different trucks in each quarter. Since analyzing intercity truck movement is our main objective, we use part of the data which comprises 18218 trips made by about 1000 trucks between 15 regions inside Denmark. The data contain information on vehicle type, commodity type, shipment size and origins and destinations of trips. All the stops a truck makes in a given day, both loaded trips and empty runs, are recorded as separate trips. One limitation of the data is that it does not have sufficient information to construct trip chains or truck tours (sub-section 5.2. discusses the implications of this limitation). But the level of disaggregation in the data is rare and it provides a significant advantage in studying capacity utilization.

Table 1 presents descriptive statistics of the main variables interest. The average load factor over the entire period for all trips is about 36 % compared to an average of 52 % for loaded trips only (i.e. excluding empty trips). The average trip distance is about 120 km. Classification of the trips to loaded and empty trips reveals that on average loaded trips (136 km) are longer than empty trips (84 km) conforming to the general assumption that carriers tend to minimize empty runs. The average carrying capacity provided in

\footnote{A notable exception is a study by Holguín-Veras (2002) which shows that trucking carriers allocate vehicles to specific segments of operation based on a joint decision process of vehicle selection and shipment size.}

\footnote{As table 1 reveals the minimum capacity is lower than 6 tonnes for the actual data. The current maximum legal carrying capacity is 52 tonnes.}

\footnote{The statistics are weighted by their occurrence (i.e. trips). Applying the expansion factor given by Denmark’s statistical authority didn’t result in significant differences in means of the main variables.}
all the trips is about 28 tons. It is also interesting to note that on average Danish carriers operate young trucks of about 4 years of age.
### Table 1: Descriptive statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Obs</th>
<th>Means</th>
<th>S. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Load factor (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole sample</td>
<td>18218</td>
<td>36.02</td>
<td>36.13</td>
<td>0</td>
<td>246</td>
</tr>
<tr>
<td>Loaded trips</td>
<td>12655</td>
<td>51.84</td>
<td>32.5</td>
<td>0.0024</td>
<td>246</td>
</tr>
<tr>
<td>Rigid truck</td>
<td>5196</td>
<td>40.01</td>
<td>42</td>
<td>0</td>
<td>246</td>
</tr>
<tr>
<td>Semi-trailer truck</td>
<td>4359</td>
<td>44.23</td>
<td>41.63</td>
<td>0</td>
<td>144</td>
</tr>
<tr>
<td>Articulated truck</td>
<td>8663</td>
<td>29.44</td>
<td>27</td>
<td>0</td>
<td>125</td>
</tr>
<tr>
<td><strong>Trip distance (km)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole sample</td>
<td>18218</td>
<td>120.12</td>
<td>99.74</td>
<td>1</td>
<td>599</td>
</tr>
<tr>
<td>Loaded trips</td>
<td>12655</td>
<td>135.85</td>
<td>108.41</td>
<td>1</td>
<td>599</td>
</tr>
<tr>
<td>Empty trips</td>
<td>5563</td>
<td>84.31</td>
<td>63.2</td>
<td>1</td>
<td>416</td>
</tr>
<tr>
<td>Rigid truck</td>
<td>5196</td>
<td>93.31</td>
<td>93.44</td>
<td>1</td>
<td>583</td>
</tr>
<tr>
<td>Semi-trailer truck</td>
<td>4359</td>
<td>113.2</td>
<td>88.77</td>
<td>1</td>
<td>576</td>
</tr>
<tr>
<td>Articulated truck</td>
<td>8663</td>
<td>139.66</td>
<td>104.31</td>
<td>2</td>
<td>599</td>
</tr>
<tr>
<td><strong>Capacity (tons)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole sample</td>
<td>18218</td>
<td>28.18</td>
<td>13.04</td>
<td>2.45</td>
<td>52.8</td>
</tr>
<tr>
<td>Rigid truck</td>
<td>5196</td>
<td>11.43</td>
<td>4.21</td>
<td>2.45</td>
<td>21.5</td>
</tr>
<tr>
<td>Semi-trailer truck</td>
<td>4359</td>
<td>30.53</td>
<td>5.24</td>
<td>8.6</td>
<td>49.15</td>
</tr>
<tr>
<td>Articulated truck</td>
<td>8663</td>
<td>37.05</td>
<td>9.13</td>
<td>11.7</td>
<td>52.8</td>
</tr>
<tr>
<td><strong>Cargo weight (tons)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole sample</td>
<td>18218</td>
<td>9.76</td>
<td>10.82</td>
<td>0</td>
<td>49</td>
</tr>
<tr>
<td>Rigid truck</td>
<td>5196</td>
<td>6.87</td>
<td>5.48</td>
<td>0.1</td>
<td>33.16</td>
</tr>
<tr>
<td>Semi-trailer truck</td>
<td>4359</td>
<td>20.12</td>
<td>10.45</td>
<td>0.1</td>
<td>49</td>
</tr>
<tr>
<td>Articulated truck</td>
<td>8663</td>
<td>15.27</td>
<td>10.2</td>
<td>0.1</td>
<td>43.7</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole sample</td>
<td>18218</td>
<td>3.76</td>
<td>3.5</td>
<td>0</td>
<td>22</td>
</tr>
</tbody>
</table>


Note: All the statistics are calculated at a trip level. Statistics for the shipment size variable are based only on loaded trips. The maximum values for the load factor are larger than 100% because of overloaded trucks.

Table 2 shows that the majority of the trips, 86%, were made by for-hire carriers while the remaining 14% were made by own-account shippers. There are three vehicle types namely: rigid trucks, semi-trailers and articulated trucks comprising 28%, 24% and 48% of the trips made in the sample period, respectively. The share of loaded, empty and fully loaded\(^7\) trips is 69%, 30% and 5%, respectively. Based on the evaluation

\(^7\)Fully loaded trips are trips with a load factor of 100% or above for overloaded trucks.
of drivers about 6% of trips were undertaken carrying voluminous (low density) cargo. Finally, there are 28 commodity groups in the dataset based on the Danish standard classification of road freight transportation goods. Three groups namely food (15%), construction (21%) and general cargo (19%) constitute the majority of the loaded trips.

Table 2: Distribution of Categorical Variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Share (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Carrier type</strong></td>
<td></td>
</tr>
<tr>
<td>For-hire</td>
<td>85.75</td>
</tr>
<tr>
<td>Own-account</td>
<td>14.25</td>
</tr>
<tr>
<td><strong>Trip type</strong></td>
<td></td>
</tr>
<tr>
<td>Loaded</td>
<td>69.48</td>
</tr>
<tr>
<td>Empty</td>
<td>30.52</td>
</tr>
<tr>
<td>Fully Loaded*</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>Truck type</strong></td>
<td></td>
</tr>
<tr>
<td>Rigid</td>
<td>28.52</td>
</tr>
<tr>
<td>Semi-trailer</td>
<td>23.93</td>
</tr>
<tr>
<td>Articulated</td>
<td>47.55</td>
</tr>
<tr>
<td><strong>Voluminous Cargo</strong></td>
<td>5.54</td>
</tr>
</tbody>
</table>

Source: The Danish heavy vehicles trip diary, 2006 and 2007

* Share of trips with load factor of 100 % and above

** Share of voluminous cargo (low density) cargo is based on the evaluation of drivers

4 The econometric framework

4.1 Joint estimation

We propose an econometric framework where the decision to run empty or loaded during a given trip, referred to as the market access decision hereafter, and the load factor are jointly estimated. Carriers are assumed to make continual market access decisions with an underlying objective of minimizing empty running, whereas in accessed markets (i.e. during a loaded trip), they aim at maximizing the load factor. We hypothesize that in a competitive environment, the market access decision depends on variables pertaining to characteristics of carriers, haul and truck. By and large similar variables determine the load factor with the exception of few that we use as an exclusion restriction (these are variables which affect the probability of market access but not the load factor). Accordingly, the joint estimation proceeds with a Heckman (1979) sample selection type model with a structural model for the load factor and a reduced form probit model for the market access decision.
The load factor is given by the following equation:

\[ Y_1 = \beta_1 X_1 + u_1 \]  

(1)

where \( X_1 \) is a vector of explanatory variables and \( u_1 \) is a residual term. \( Y_1 \) is the load factor, whose observability is conditional on the following market access equation:

\[ Y_2 = \begin{cases} 1 & \text{if } \sigma_2 X_2 + v_2 \geq 0 \\ 0 & \text{otherwise} \end{cases} \]  

(2)

where \( Y_2 = 1 \) if a truck is loaded, \( X_2 \) contains all variables in \( X_1 \) and additional variables for identification (exclusion restriction) and \( v_2 \) is a residual term. \( X_2 \) is always observed, regardless of \( Y_2 \). The following assumptions are made for estimation (Wooldridge, 2000 p. 562):

a. \((X_2, Y_2)\) are always observed, but \( Y_1 \) is only observed when \( Y_2 = 1 \);

b. \((u_1, v_2)\) are independent of \( X_2 \) with zero means (\( X_2 \) is exogenous in the population)

c. \( u_1 \sim \text{normal} (0, 1) \) and \( v_2 \sim \text{normal} (0, 1) \)

d. \( E(u_1 \mid v_2) = \gamma_2 v_2 \) (residuals may be correlated; e.g. bivariate normality).

Assumption 'a' is slightly different in our case since \( Y_1 \) - the load factor- is always observed, but it takes a value equal to zero when \( Y_2 = 0 \). The difference has implications on how we calculate partial effects (see section 3.3.). Assumptions ‘c’ and ‘d’ imply that the covariance between the two error terms is \( \gamma_2 \). A sample selection problem arises if \( \gamma_2 \) is not statistically different from zero suggesting that the residual in the load factor equation is correlated with the residual in the market access equation.

Our methodology is usually referred to as a discrete-continuous modeling in the transport economics literature. It has been widely applied in many passenger transportation problems (see Mannering and Hensher, 1987 for an excellent early review). Holguin-Veras (2002) applies a somewhat similar methodology to study vehicle selection process and shipment size, assuming that they are simultaneously determined. In the present study, the joint estimation is selected because we are interested both in the determinants of the load factor and the market access decision (empty/loaded movement decision). Modeling the load factor exclusively based on loaded trips (dropping empty trips), introduces a sample selection problem. This is because an empty trip does not necessarily imply that carriers were not offered a loading opportunity. They may have been offered the opportunity, but decided to run an empty truck due to higher costs compared to net revenue (Beilock and Kilmer, 1986). Therefore, a joint estimation is needed to model the load factor to understand its determinants at a population level.\(^8\)

Barla et al (2010) estimate equation 1 using a multinomial ordered probit model to find the effect of information technology capability of a truck on the load factor. In their specification, the load factor is

\(^8\)A joint estimation can also be warranted due to omitted variables in our sample such as information technology capability that lead to a correlation in the error terms.
classified into five discrete groups, and both loaded and empty trips are included in $Y_1$. Doing so implies that a single mechanism determines the load factor and the probability of being loaded. However, it is possible that trucks with a high probability of being loaded also tend to have a high load factor, and vice-versa. The econometric specification should, therefore, account for and explain why some trucks have a higher load factor than others jointly with why some trucks tend to move loaded more frequently than others. In the present study, we aim to give such an explanation by estimating a reduced form probit model for the market access decision jointly with the load factor. The following subsections present an outline of the relationship between potential explanatory variables and the two dependent variables.

### 4.1.1 Determinants of the load factor

The load factor is measured as the percentage share of a truck’s loading capacity that is filled with a cargo. Its level is constrained both by the weight and volume (density) of the cargo. Ideally, both constraints should be taken into account to use the load factor as a capacity utilization measure. Unfortunately, our data does not contain an exact measure of volume, but we instead control for the effect of density by including commodity class dummy variables and a dummy variable which indicates whether a cargo is voluminous or not based on the evaluation of truck drivers. The following formula is used to calculate the load factor based on only the cargo weight constraint:

$$LF(\%) = \frac{CW}{MC} \times 100$$  \hspace{1cm} (3)

where $LF$, $CW$ and $MC$ stand for the load factor, cargo weight and the maximum legal carrying capacity of a truck, respectively. Carriers generally want to maximize the load factor for their trucks for two main reasons. First, it is conceivable that profit margins depend on how often carriers can have their truck filled to its potential. Second, the load factor is one of the key determinants of energy efficiency since a high load factor implies a higher level of ton-kilometer (output) for a given vehicle-kilometer (input). Energy requirement, however, increases less than proportionally with the load factor over a distance. As indicated by Baral et al. (2010), a fully loaded truck consumes only about 20% more fuel compared to an empty truck. Hence, assuming that carriers maximize the load factor, we re-specify equation 1 for estimation as follows:

$$y_{1i} = \beta_1 x_{1i} + u_{1i}$$  \hspace{1cm} (4)

All the variables in equation 4 are defined as in 4.1, and they are measured at a trip level, i.e. $x_{1i}$ include explanatory variables pertaining to characteristics of haul, truck and carrier. We control for two haul characteristics, namely the type of commodity carried and trip distance. Using dummy variables that indicate the commodity type of carried cargo, we capture the effect of density on the load factor. Doing so reduces heterogeneity biases that may result from comparing heavy and dense cargo to light and low density cargo that cubes-out a vehicle space before the maximum carrying capacity (in terms of weight) is reached.
In addition to showing the effect of density, the commodity dummies reveal, to a reasonable degree, shipper characteristics and its effect on the load factor. To see the pure effect of density on the load factor, we include a dummy variable which indicates whether a cargo is voluminous or not based on the evaluation of drivers, and it is expected to have a negative effect on the load factor. As for the effect of trip distance, the load factor tends to be higher for trucks hauling cargo over long distance because of the high opportunity cost of running trucks partially filled (Barla et al, 2010). Distance is, therefore, expected to have a positive effect on the load factor.

Truck characteristics such as loading capacity may affect the load factor in two opposite directions. On the one hand, carriers may find filling a small truck easy since it is maneuverable to aggregate loads from different shippers (Holguin-veras, 2002). Carriers may also follow an optimization strategy of filling the largest available load to the smallest available vehicle capacity. Small trucks, therefore, may tend to have a higher load factor compared to larger trucks (Fowkes, 2007). On the other hand, just-in-time inventory strategies of shippers may prompt carriers to use smaller trucks that are suitable for small shipments that may result in lower load factor. It is also possible that large trucks are filled to their capacity more often than small trucks as carriers try to avoid the relatively high opportunity cost of running partially filled large trucks and hence reversing the effect of capacity on the load factor.

The likely cause of the opposing effects of capacity on the load factor is the non-linearity of trucks’ carrying capacity. While the range of a cargo size for shipment is more-or-less continuous, vehicle capacity is rather discontinuous, and trucks are classified into different vehicle classes of fixed carrying capacity. To uncover the two opposing effects of a truck’s size on the load factor, we include two variables: capacity (measured as the maximum legally allowable carrying capacity) and its squared terms. We expect a negative sign for the former reflecting that a small truck can easily be filled fully, and a positive sign for the latter reflecting the claim that carriers dislike to operate a large truck that is filled partially. Finally, we expect that for-hire carriers tend to have a higher load factor compared to own-account shippers since they usually have more incentive and flexibility to find complementary demands compared to own-account carriers (Hubbard, 2003; Barla et al., 2010).

4.1.2 Determinants of market access

The market access decision, defined as the probability of being loaded or empty, is captured by equation 2. We estimate the equation by the following binary probit model:

\[ y_{2i} = \sigma_2 x_{2i} + \nu_{2i} \]  

(5)

All the variables in equation 5 are defined as in 4.1, and they are measured at a trip level, i.e. \( x_{2i} \) include observable carrier, haul and truck characteristics such as carrier type, trip distance, freight movement balance between regions and a truck’s carrying capacity and age. Looking further at the effect of these variables, for-hire carriers are expected to have a better market access than own-account carriers because they have
specialized staff engaged in finding complementary demand (Hubbard, 2003). In contrast, own-account carriers may have a higher opportunity cost of market access because they often have prior commitments to proceed to other destinations to haul their firms’ good (Wilson and Beilock, 1994). Therefore, we expect a positive sign for a dummy variable that takes a value equal to one if a truck is owned by a for-hire carrier or zero otherwise.

We included two variables, namely trip distance and a dummy variable for freight movement balance, to capture the effect of haul characteristics on the market access decision. It is usually the case that the longer the trip distance, the higher will be the probability of market access since operating an empty truck over a long distance is costly. Between nearby locations, however, trucks may be seen running empty for repositioning or refueling purposes. It is also possible that finding complementary demand for short distance trips can be difficult, increasing the chances of empty runs (Barla et al., 2010; Wilson and Dooly, 1993). Thus, trip distance is expected to influence the probability of being loaded positively.

The freight movement balance between regions is also an important determinant of the probability of getting a load. The reason lies in market access expectations of carriers. For instance, it is likely that carriers get a return load more often from a net exporting region than from a net importing region. Conversely, the probability of empty running to a net importing region is lower since carriers usually make a loaded trip to the region because of slimmer chances of getting a return load (see Beilock and Kilmer, 1986 for an interesting discussion). To capture the effect of freight balance, we include a dummy variable which equals one if a trip is made toward a net importing region and zero otherwise. And its effect on market access is expected to be positive.

Finally, we note that more variables are required to capture market access decision fully. The main focus has been on the variables discussed above because we think that they explain most of the variation in the market access decision, at least in economic terms. There is also a data limitation. For instance, variables that capture market conditions that lead to trip generation are missing in our analysis. Controlling for whether a truck is tied up to serve a specific shipper (as in Beilock and Kilmer, 1986) during the survey period may explain why a carrier forgoes loading opportunities. Information technology capability of trucks and additional carrier specific variables (such as fleet size and membership of online load sharing arrangements) should have also been controlled for. Exclusion of these and other variables matters, however, in so far as they are correlated with the variables we control for. At this point we assume that there is no such a correlation and proceed with the proposed joint estimation. In section 4.2 we explain potential problems of this assumption and provide alternative models to address them. Table 3 presents list of the main explanatory variables.

---

The regions in our dataset are classified as net exporting and net importing regions based on the size of freight movement in our dataset (a physical measure) rather than a trade flow expressed in monetary unit.
Table 3: List of variables (var.) used in the empirical analysis

<table>
<thead>
<tr>
<th>Category</th>
<th>Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent var.</strong></td>
<td>LF</td>
<td>Load factor</td>
</tr>
<tr>
<td></td>
<td>Loaded</td>
<td>Dummy var. equals 1 if LF &gt; 0 and 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>For-hire</td>
<td>Dummy var. equals 1 if a truck is owned by a for-hire carrier and 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>Rigid</td>
<td>Dummy var. equals 1 if a truck is rigid and 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>Semi-Trailer</td>
<td>Dummy var. equals 1 if a truck is a semi-trailer and 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>Artic</td>
<td>Dummy var. equals 1 if a truck is articulated and 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>Age of a truck</td>
</tr>
<tr>
<td></td>
<td>Capacity</td>
<td>Maximum legal carrying capacity of a truck</td>
</tr>
<tr>
<td></td>
<td>Distance</td>
<td>Distance between origin and destination of a trip</td>
</tr>
<tr>
<td></td>
<td>Trip to net-import region</td>
<td>Dummy var. equals 1 if a trip is made towards a net importing region</td>
</tr>
<tr>
<td></td>
<td>Commodity</td>
<td>set of dummy var. showing the commodity class of a cargo</td>
</tr>
</tbody>
</table>

The proposed joint estimation requires one or more variables that affect the market access decision but not the load factor. We use the freight balance dummy variables and the age of a truck as exclusion restrictions in the market access model (eq. 5). We think that a difference in freight flow between regions is more likely to affect how often a truck is loaded, but not how intensively it is used during a loaded trip. Concerning the validity of a truck’s age as an exclusion restriction, a correlation between a truck’s vintage and the load factor most likely reflects variations in the probability of market access, but not variations in the level of capacity utilization during a loaded trip. This is a plausible assertion given the general trend that carriers tend to use their younger trucks more often to rest older trucks, especially when faced with excess capacity (Hubbard, 2003). There are two pieces of evidence in the data that support our assertion. First, the age of a truck has a significant (at one per cent) and negative correlation both with the number of days a truck is used and the number of kilometers it is driven with a load, implying preference for younger trucks. Second, we find a rather small but significant (at one per cent) and positive correlation between the load factor and age. The finding again implies that old trucks are loaded as full as new ones, if not more, during a loaded trip.
5 Results

5.1 Main results

Table 4 presents results from four different specifications based on full information maximum likelihood estimation (FIML).\textsuperscript{10} All the continuous variables are in levels. We use two exclusion restrictions in the load factor equation (eq. 4); in each specification ‘age’ and the freight movement balance indicator variable (‘trip to net import’) are included in equation 5 and excluded from equation 4. In line with our expectation, both variables have significant (at 1 percent) negative and positive effects on the probability of markets access, respectively. The appropriateness of the joint estimation is confirmed by the significance of $\rho$, the correlation coefficient between the residual terms in the two equations. The result implies that the subsample of loaded trips is not random, and hence the joint estimation corrects for potential sample selection bias (see the last four paragraphs of this sub-section for more explanation).

-Table 4 about here-

In the first specification (column one), though few variables are controlled for it is interesting to see that all the main variables have the expected sign and statistically significant effects on the load factor and the market access decision. The second specification controls for the commodity group of a cargo to capture how the load factor differs between the various shippers served by carriers.\textsuperscript{11} It is important to note that the commodity class dummy variables may partly reveal the effect of density since there are size differences between cargoes and their packaging requirements. To disentangle the dual effect of the commodity type, the third specification directly controls for the effect of density by a dummy variable which indicates whether the carried cargo is voluminous or not. Accordingly, we find that being a voluminous cargo has a significant and negative effect on the load factor. Further, the slightly lower effects of the commodity dummy variables in the third specification (compared to the second) capture how serving different shippers affects the load factor.

A truck’s size, measured by capacity (in tonnes), appears to have a negative effect both on the load factor and the probability of market access for the first three specifications. The coefficient of capacity-squared in the fourth specification, however, shows that the effect of capacity depends on the size of trucks, and as such for larger trucks capacity seems to have a significant positive effect on the load factor.\textsuperscript{12} The result shows

\textsuperscript{10}A two step model similar to Heckman (1979) resulted in almost identical results with a significant and negative sign for the inverse Mill’s ratio. But we opted for the FIML results which are proven to give more efficient estimates (Puhani, 2000).

\textsuperscript{11}We cannot directly control for commodity type in the probit model because we are interested in the probability of being loaded or carrying a commodity. We instead tried to include commodity information (and hence shippers characteristics) indirectly using a dummy variable which indicates the typical commodity carried by each truck. But the attempt gave counter intuitive results.

\textsuperscript{12}Using different sets of methodology and dataset to the ones used here, a study by the American Transportation Research Institute (2008) shows that the overall environmental efficiency in terms of freight transported per unit of CO2 emitted is better for heavy vehicles. Similarly, McKinnon (2005) concludes that the UK government’s decision to increase the maximum truck weight to 44 tons in 2001 led to a “significant economic and environmental benefits”.
that the effect of truck size on utilization may be non-linear. We also note that inclusion of capacity-squared leads to an insignificant coefficient for capacity in the markets access equation. Nonetheless, the effect of truck size is in line with our hypothesis that as truck size gets bigger, utilization deteriorates up to a point and then improves.

In all the specifications, distance (as expected) is shown to have a positive and significant (at 1 percent) effect both on the load factor and the probability of market access. Its point estimates are, however, rather small. Finally, the co-efficient of ‘for-hire’ or not shows that being a for-hire carrier does not seem to have a significant effect on the market access decision, but it has a positive and significant (at 1 percent) effect on the load factor. The result confirms the hypothesis that for-hire carriers are more capable in aggregating loads during a given trip compared to own-account carriers (Hubbard, 2003).

The negative sign for rho may appear anomalous since most studies based on sample selection models get a positive sign for it. In our context, a positive sign may also sound more ‘plausible’. The reason is that the unmeasured effects that increase the chances of market access are also likely to increase the load factor. A negative correlation, however, implies that a truck that carries a load, when it is predicted to be unlikely to be loaded on the basis of the market access equation has a lower load factor than would be predicted from the load factor equation on the basis of the measured characteristics. A negative correlation is not uncommon in the literature, and we give the following three explanations.

First, it is important to note that there is no prior reason to expect a positive relationship between the two error terms. A theoretical paper by Ermisch and Wright (1994) shows that a negative correlation can arise if the variance of the error term in the structural equation (in our case the load factor equation) is less than the variance of the error term for the latent variable in the reduced form probit equation (in our case the market access equation). That is, given the exogenous variables in the load factor and the market access equations, the load factor exhibits less dispersion than the latent variable. As indicated in section 3, our estimates are based on the assumption that a higher level of load factor leads to a larger profit. In fact, implicit in the market access probit equation is an optimization process where carriers access a market if net profit is greater than or equal to zero. Therefore, it is possible that the dispersion of net profit, the latent variable, is greater than the dispersion in the load factor, resulting in a negative correlation between the two residual terms.

Second, the fact that we restrict our sample to intercity trips may introduce selectivity problem. For instance, the correlation between the residuals might be positive in the whole sample that includes trips within regions), but it might be negative among the sub-population of intercity trips because of the nature of the selectivity process. Third, a more plausible explanation for the negative correlation is related to carriers’ expectations. For a truck in a backhaul trip it is usually difficult to get a return load. A carrier may, therefore, choose to carry a small load (implying lower load factor) instead of running empty if freight rates cover market access costs (implying less empty running). A negative correlation can also occur because of carriers’ tendency to have a higher level of load factor on trucks for which there is no anticipated return.
load compared to a truck expected to be loaded both during a front-haul and backhaul directions. Similarly, carriers may forego a loading opportunity in outbound direction if they anticipate a backload leading to a negative correlation between the probability of market access and the load factor.

5.2 Robustness checks

To check whether our main results in table 4 are robust we tried several alternative specifications. A censored tobit model (not presented here), apart from having larger coefficient estimates, gave a comparable result both in terms of significance and sign. But our attempt to include trip distance as a categorical variable defined as short, medium and long did not give sensible results. Inclusion of a quadratic term of trip distance in the main specification also resulted in convergence problems. Table 5 presents results based on four alternative models using full information maximum likelihood estimation (FIML). All the continuous variables are in levels. To capture differences in capacity utilization between the three vehicle classes in the data, dummy variables that indicate the type of a truck is included in all the models.\(^\text{13}\) In the load factor equation, articulated trucks appear to be utilized less efficiently compared to rigid trucks. The same result holds for semi-trailers only under model one. In the market access equation, however, we find no statistically significant difference in utilization between the trucks. Further, inclusion of interaction effects between the vehicle class dummy variables and the capacity variable in models 3 and 4 reveals that the effect of capacity does not depend on truck type. The signs and significance of the main variables of interest, such as carrier type, trip distance capacity and rho are very similar to those in table 4, confirming the robustness of the overall estimates.

\(^\text{13}\)Separate regressions for each vehicle class also gave very comparable result.
observed time invariant variables in $X_1$ and $X_2$ (such as, age, capacity, and carrier type). This is obviously an uninteresting outcome given the relevance of the variables to our analysis. Besides, the fixed effects model is consistent and asymptotically normal for short panels only (Wooldridge, 2002). In our data, however, some trucks make up to 99 trips, implying that the panel structure in the data is rather long.

We note that our analysis may suffer from two potential specification problems, namely omitted variables and simultaneity bias. We first discuss implications of the omitted variables. The econometric models are based on the assumptions that all trips made by a truck are independent. However, some of the trips might be part of a trip-chain or a tour. If so, both the market access decision and the load factor may not be determined at an individual trip level (as it is assumed here); hence, the interdependence between trips within a trip-chain should have been controlled for. Since we use only inter-city trips, however, the interdependency problem between the trips is minimized. Finally, we note that capacity utilization may depend also on missing variables such as fleet size.\footnote{Data on ICT capability of a truck, membership in online load sharing arrangement and carrier size could also have added more realism to our modeling.} A small carrier (in terms of fleet size) may appear less efficient if it is forced to use a rather big truck for a small load compared to a larger carrier that matches its truck capacity with available load. But we think this might not be always the case. Regardless of their fleet size, carriers may sometimes choose efficient truck-haul combinations in a front-haul trip, but not during a backhaul trip.

Concerning the second specification problem, the maximum carrying capacity of a truck might be simultaneously determined with how much to carry and hence with the load factor. The simultaneity implies that capacity, a right hand side variable, can potentially be endogenous in the load factor equation. In fact, since the load factor measures the percentage share of a truck’s carrying capacity that is filled with a load, capacity appears as a denominator on the right hand side of equation 4. Addressing the simultaneity problem calls for an explicit consideration of carriers’ vehicle selection process. A variant of a nested logit model with three layers, where the vehicle selection process, the market access decision (empty/loaded movement decision) and the load factor are jointly modeled can be used to handle the problem. It is a promising way for future studies, but it is beyond the scope of the present study. Finally, our results, though interesting and in line with our main hypotheses, should be interpreted carefully with the above caveats in mind.

6 Conclusions

We have explained the underlying determinants of two aspects of capacity utilization in road freight transport, the extent of empty running and the load factor. Detailed information about a truck’s operation at a trip level has enabled us to model the empty/loaded movement decision and the load factor jointly to account for a sample selection problem. The results indicate that the main explanatory variables have the expected signs and statistically significant effects. Trip distance is shown to have a positive effect both on the load
factor and the probability of market access. Further, though being a for-hire carrier or not does not appear to have a significant effect on the market access decision, it has a positive and significant effect on the load factor. The result confirms the premise that for-hire carriers are more capable in aggregating loads from different shippers compared to own-account carriers. Truck size, captured by the maximum allowable carrying capacity, appears to have a negative effect both on the load factor and market access. A significant and positive coefficient for its squared term shows that the effect of carrying capacity is non-linear; as such for larger trucks the market access probability and the load factor are higher.
References


Table 4: Estimates for the load factor and the market access equations

<table>
<thead>
<tr>
<th>Variables</th>
<th>[1]</th>
<th>[2]</th>
<th>[3]</th>
<th>[4]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.555</td>
<td>27.85</td>
<td>0.589</td>
<td>8.33</td>
</tr>
<tr>
<td>For-hire</td>
<td>0.048</td>
<td>5.60</td>
<td>0.076</td>
<td>9.86</td>
</tr>
<tr>
<td>Own-account</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>0.0002</td>
<td>3.80</td>
<td>0.0003</td>
<td>8.46</td>
</tr>
<tr>
<td>Capacity</td>
<td>-0.004</td>
<td>-17.17</td>
<td>-0.006</td>
<td>-28.30</td>
</tr>
<tr>
<td>Commodity dummies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Live animals</td>
<td>-0.031</td>
<td>-0.43</td>
<td>-0.039</td>
<td>-0.55</td>
</tr>
<tr>
<td>Cereals</td>
<td>0.347</td>
<td>4.89</td>
<td>0.343</td>
<td>4.87</td>
</tr>
<tr>
<td>Potatoes</td>
<td>-0.074</td>
<td>-1.04</td>
<td>-0.062</td>
<td>-0.88</td>
</tr>
<tr>
<td>Sugar beets</td>
<td>0.238</td>
<td>2.40</td>
<td>0.266</td>
<td>2.69</td>
</tr>
<tr>
<td>Wood</td>
<td>0.057</td>
<td>0.82</td>
<td>0.062</td>
<td>0.90</td>
</tr>
<tr>
<td>Skin and textile</td>
<td>-0.265</td>
<td>-3.26</td>
<td>-0.268</td>
<td>-3.32</td>
</tr>
<tr>
<td>Food</td>
<td>-0.039</td>
<td>-0.56</td>
<td>-0.041</td>
<td>-0.59</td>
</tr>
<tr>
<td>Feed and fodder</td>
<td>0.269</td>
<td>3.84</td>
<td>0.265</td>
<td>3.79</td>
</tr>
<tr>
<td>Fatty substance</td>
<td>0.240</td>
<td>3.19</td>
<td>0.232</td>
<td>3.11</td>
</tr>
<tr>
<td>Coal</td>
<td>0.217</td>
<td>2.51</td>
<td>0.210</td>
<td>2.44</td>
</tr>
<tr>
<td>Crude oil</td>
<td>-0.016</td>
<td>-1.61</td>
<td>-0.016</td>
<td>-1.68</td>
</tr>
<tr>
<td>Petrol and minl. oil</td>
<td>0.121</td>
<td>1.70</td>
<td>0.119</td>
<td>1.67</td>
</tr>
<tr>
<td>Iron ore &amp; scrap</td>
<td>0.138</td>
<td>1.89</td>
<td>0.132</td>
<td>1.82</td>
</tr>
<tr>
<td>Ore &amp; Scrp of metal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-prod. Metal</td>
<td>-0.050</td>
<td>-0.69</td>
<td>-0.055</td>
<td>0.07</td>
</tr>
<tr>
<td>Gravel,sand, salt,stone,dirt</td>
<td>0.227</td>
<td>3.26</td>
<td>0.224</td>
<td>0.07</td>
</tr>
<tr>
<td>Cement,brick, L.st.</td>
<td>0.140</td>
<td>2.01</td>
<td>0.141</td>
<td>0.07</td>
</tr>
<tr>
<td>Fertilisers</td>
<td>0.298</td>
<td>4.01</td>
<td>0.292</td>
<td>0.07</td>
</tr>
<tr>
<td>Tar and asphalt</td>
<td>0.166</td>
<td>2.26</td>
<td>0.160</td>
<td>0.07</td>
</tr>
<tr>
<td>Chemical products</td>
<td>-0.066</td>
<td>-0.91</td>
<td>-0.061</td>
<td>0.07</td>
</tr>
<tr>
<td>Cellulose &amp; waste of paper</td>
<td>-0.112</td>
<td>-1.44</td>
<td>-0.112</td>
<td>0.08</td>
</tr>
<tr>
<td>Machinery</td>
<td>-0.050</td>
<td>-0.72</td>
<td>-0.050</td>
<td>0.07</td>
</tr>
<tr>
<td>Metal products</td>
<td>-0.107</td>
<td>-1.51</td>
<td>-0.103</td>
<td>0.07</td>
</tr>
<tr>
<td>Glass and Ceramics</td>
<td>-0.050</td>
<td>-0.65</td>
<td>-0.053</td>
<td>0.08</td>
</tr>
<tr>
<td>Furniture,paper, cl.</td>
<td>-0.169</td>
<td>-2.42</td>
<td>-0.149</td>
<td>0.07</td>
</tr>
<tr>
<td>house furniture</td>
<td>-0.205</td>
<td>-2.61</td>
<td>-0.178</td>
<td>0.08</td>
</tr>
<tr>
<td>General cargo</td>
<td>-0.134</td>
<td>-1.93</td>
<td>-0.127</td>
<td>0.07</td>
</tr>
<tr>
<td>Empty continers</td>
<td>-0.372</td>
<td>-5.29</td>
<td>-0.378</td>
<td>0.07</td>
</tr>
<tr>
<td>Effect of Density</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voluminous cargo</td>
<td>-0.107</td>
<td>-11.5</td>
<td>-0.1038</td>
<td>-11.16</td>
</tr>
<tr>
<td>Quadratic term</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capacity squared</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rho</td>
<td>0.049</td>
<td>0.61</td>
<td>-0.183</td>
<td>-2.59</td>
</tr>
<tr>
<td>Sigma</td>
<td>0.321</td>
<td>152.2</td>
<td>0.279</td>
<td>112.8</td>
</tr>
</tbody>
</table>
Table 4: - continued-

<table>
<thead>
<tr>
<th>Variables</th>
<th>[1]</th>
<th>[2]</th>
<th>[3]</th>
<th>[4]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>t-stat</td>
<td>Coeff.</td>
<td>t-stat</td>
</tr>
<tr>
<td>Constant</td>
<td>0.108</td>
<td>2.72</td>
<td>0.099</td>
<td>2.54</td>
</tr>
<tr>
<td>For-hire</td>
<td>0.016</td>
<td>0.57</td>
<td>0.015</td>
<td>0.51</td>
</tr>
<tr>
<td>Own-account</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>0.004</td>
<td>32.0</td>
<td>0.004</td>
<td>32.27</td>
</tr>
<tr>
<td>Capacity</td>
<td>-0.0034</td>
<td>-4.09</td>
<td>-0.003</td>
<td>-4.08</td>
</tr>
<tr>
<td>Capacity squared</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truck age</td>
<td>-0.011</td>
<td>-3.51</td>
<td>-0.009</td>
<td>-3.21</td>
</tr>
<tr>
<td>Trip to Net-import</td>
<td>0.157</td>
<td>7.58</td>
<td>0.156</td>
<td>7.58</td>
</tr>
</tbody>
</table>

*The market access equation*

Log L: -14129, -12265, -12198, -12142
Observation: the market access equation N = 18218, the load factor equation N = 12657

Note: the reference category for the commodity dummy variables, ore and scrap metals, has a load factor equal to the sample average (51).
### Table 5: Alternative Estimates

<table>
<thead>
<tr>
<th>Variables</th>
<th>[1]</th>
<th>[2]</th>
<th>[3]</th>
<th>[4]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>t-stat</td>
<td>Coeff.</td>
<td>t-stat</td>
</tr>
<tr>
<td>Constant</td>
<td>0.514</td>
<td>7.43</td>
<td>0.559</td>
<td>7.98</td>
</tr>
<tr>
<td>For-hire</td>
<td>0.087</td>
<td>11.56</td>
<td>0.088</td>
<td>11.66</td>
</tr>
<tr>
<td>Own-account</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Distance</td>
<td>0.0004</td>
<td>9.56</td>
<td>0.0004</td>
<td>9.37</td>
</tr>
<tr>
<td>Voluminous cargo</td>
<td>-0.10</td>
<td>10.94</td>
<td>-0.099</td>
<td>10.76</td>
</tr>
<tr>
<td>Rigid Truck</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Semi Trailer Truck</td>
<td>-0.044</td>
<td>-4.57</td>
<td>0.040</td>
<td>1.23</td>
</tr>
<tr>
<td>Artic. Truck</td>
<td>-0.191</td>
<td>-17.18</td>
<td>-0.280</td>
<td>-14.00</td>
</tr>
<tr>
<td>Capacity</td>
<td>-0.001</td>
<td>-2.30</td>
<td>-0.005</td>
<td>-4.44</td>
</tr>
<tr>
<td>Rigid*Capacity</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Semi Trailer*Capacity</td>
<td>-0.0001</td>
<td>-0.09</td>
<td>-0.001</td>
<td>-0.65</td>
</tr>
<tr>
<td>Artic. * Capacity</td>
<td>0.005</td>
<td>4.51</td>
<td>0.004</td>
<td>1.72</td>
</tr>
<tr>
<td>CapacitySquared</td>
<td>0.0001</td>
<td>4.65</td>
<td>0.0003</td>
<td>0.84</td>
</tr>
<tr>
<td>Commodity dummies</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Rho</td>
<td>-0.187</td>
<td>-2.60</td>
<td>-0.178</td>
<td>-2.38</td>
</tr>
<tr>
<td>Sigma</td>
<td>0.273</td>
<td>110.3</td>
<td>0.272</td>
<td>111.27</td>
</tr>
</tbody>
</table>

### The load factor equation

Observations: market access equation $N = 18218$, load factor equation $N = 12657$
Introduction

The transportation sector is a significant polluter in terms of climate change, but also in terms of local air pollution. In EU, the transport sector’s share of GHG emissions has been growing continuously since 1990 (EEA 2007). It has become evident that current single and isolated instruments are not sufficient to curb such a trend. In order to alter this development, current policy approaches has to be reconsidered. Policy packages are recognized as a more integrated and holistic approach, which potentially can move transport policies into a more environmental sustainable direction.

However, there are still gaps in knowledge about policy packages. Previous research has pointed out that policy packages often are composed in an incidental manner and they are seldom used as intended by local transport planners (Hull 2008). An increased complexity among the actors involved can thus lead to implementation deficit resulting from lack of policy integration. In addition, few studies have analyzed and evaluated policy packages. Consequently, less is known regarding policy packages effectiveness, and identifying which combination of instruments is more likely to successfully achieve sustainable transport.

This paper will look at a case study from Bergen, which implemented a policy package to reduce local air pollution, and analyze whether it reached its intended goals. In addition, the combination of instruments employed is studied to analyze the design of policy packages. These were short term objectives, and mitigation of long term effects is another issue.

The structure of the paper is as follows. First, I will develop a framework for analyzing policy packages. Instruments are classified as direct, indirect and whether they are legislative, facilitative or informative. Then instruments effectiveness is considered in terms of single effect and combinational effects. The case study will illustrate that only a few instruments proved to be effective. I argue that the case of Bergen had low goal achievement since the instruments were to a large extent indirect and weak. However, used in combination with

---

1 This article is based on a report financed by Bergen municipality (Strand et al. 2010).
direct and strong instruments the policy package would probably attain higher effectiveness. The case study illustrates the need for a careful selection of incentives and the necessity of a profound understanding on how the various instruments interact in order to increase effectiveness of a policy package.

Method

A crucial question is how to evaluate policy packages and in terms of what. There are numerous ways of evaluating policies. Policies can be evaluated in terms of effectiveness, efficiency or legitimacy (Vedung 1998). Although all these aspects are important, the length of this paper does not allow going into detail of all criteria’s in evaluating policies. In this case, policy strategies will be evaluated in terms of effectiveness. This means that instruments will be analyzed in terms on to which degree they attain goal achievement (van der Doelen 1998). Literature regarding policy packages highlights their advantage in promoting effectiveness. Therefore it is of interest to test the extent of such effects.

Bergen municipality’s goal was to reduce the level of local air pollution. Local air pollution consists mainly of two components; nitrogen dioxide and fine particulate air pollution. In Bergen, road traffic constitutes more than 60% of emissions regarding nitrogen dioxide. Shipping and aviation contributes to about 15%. Other sources, as refuse incineration and private households, amount for the remaining 25% (Action Plan for increased air quality in Bergen 2007). Reducing road traffic is therefore a major polluter and it is necessary to reduce traffic in order to reduce local air pollution, both in short and long term. Methodologically, the instruments are tested in how they affected traffic levels during January and February 2010. Traffic surveys, available from Norwegian road authorities, are used to count total number of cars passing Bergen each day in this period. Instruments are expected to have an affect shortly after it is initiated. Interviews were also conducted with representatives from Bergen municipality and the Norwegian Road authorities.

2 Fine particulate air pollution has a different composition. Private households contribute to about 65-70 %, while traffic contributes to about 15 %. Consequently, instruments should focus on reducing traffic if the aim is reduced levels of nitrogen dioxide. Reducing fine particulate air pollution has a different target group; private households. Consequently, this paper analyzes how policy instruments were effective in reducing traffic to reduce the level of nitrogen dioxide.
Policy-package

A policy-package is defined by Givoni et. al (2011:3):

(...)As a combination of individual policy instruments, aimed at addressing one or more policy goals. The package is created in order to improve impacts of the individual policy instruments, minimize possible negative side effects, and/or facilitate interventions implementation and acceptability.

This definition stresses that it is necessary that a policy-package develop in a coherent manner. They also note that a policy package is often an ad-hoc process where decisions often are taken in isolation from each other (ibid). By using the above definition, it is vital that the package is created to improve the impacts of the individual policy instruments and aimed at addressing one or more policy goals. The underlying rational is that it is necessary and beneficial to include a policy package in order to optimally reach objectives. There are no “silver bullet” instruments given the multidimensionality of transport and mobility (Givoni et al. 2010). Transport policy is a complex matter and therefore a combination of instruments is often necessary in order to reach an objective (OPTIC 2010).

Several recent studies have recently been conducted in this field. More specifically Marshall and Banister (2000) argue that it is necessary to connect travel reduction strategies into a policy package in order to reach policy objectives. Sørensen et al. (2011) have identified success factors and barriers to implement policy packages in transport. May and Crass (2007) have also demonstrated that measures is likely to meet barriers in terms of implementation. Banister (2008) argues that policy packages are a key element to achieve public acceptability of sustainable transport. Banister et al. (2000) have also identified key aspects in a policy package. OPTIC (2010) have studied unintended effects and analyzed how policy packages can mitigate such effects. May and Roberts (1995) have explored how packages can creates synergies in terms of complimentarity, financial support and public acceptability.

Analyzing policy-packages

Researchers have called for a more integrated and holistic policy approach in the transport sector. However, few studies have been conducted to evaluate policy packages or study their effectiveness. There is a need of a deeper understanding on how policy packages should be developed and there is limited evidence on their success. The next section will, therefore, lay out a framework, which can be
used to analyze such an aspect and look into how policy packages can be effective.

**Direct and indirect instruments**

Table 1 illustrates a framework which can be used in order to analyze policy packages effect and can provide information on what instruments is necessary to reach an objective. The left column illustrates a given (combination of) instruments. Instruments can address different aspects of a problem and can be classified into two broad categories; direct and indirect instruments. Direct instruments are aimed at the core of a problem. This means that it seeks to directly to alter or influence an issue. For instance a congestion tax is an instrument which directly affects road traffic. Indirect instruments work with other variables and can facilitate change of behavior by making alternatives more favorable (Strand et al. 2010). Strengthening public transport in order to reduce the number of cars might be an example. The rationale behind this classification is to highlight the precision of the various instruments. In theory, it is expected a link in effectiveness between direct and indirect instruments. Direct instruments aims at the core of a problem, while indirect instruments seek to divert or make alternatives more attractive. It is therefore less precise and one could therefore expect that an (direct) economic instrument will have greater impact on travel behavior compared to an indirect instrument. Literature regarding policy instruments effectiveness in reducing transport might illustrate such a point (for instance Marshall and Banister 2000). Restrictive instruments directed at reducing transport are seen as more efficient than for instance strengthening public transport.

**Legislative, facilitative incentive and information**

Instruments can again be divided into three broad subcategories; legislative, facilitative and information. This is very much similar to Vedung’s (1998) classification into regulation, economic means and information. Of course, economic incentives are a highly relevant category. Implementing road pricing schemes or reducing public transport fares are just two examples, but taxes and subsidies can also be important. However, in the case of Bergen, the use of economic instruments was not extensively employed. Instead, incentives as facilitating instruments were more relevant, for instance increasing the public transport supply. Consequently, there are more useful to use such terms in this paper.

Directive or legislative refers to general rules and standards. Facilitative instruments refer to instances which aimed at strengthening public transport or
create incentives to share cars. HOV-lanes can be an example of the latter. The last category includes instruments intended to educate or spread norms and ideas to the public. This is the “weakest” form of instruments, since authorities rely on information in order to bring about change. Citizens can choose by themselves whether to follow recommendations or campaigns aimed at changing behavior or reach an objective. There are probably differences in terms of efficiency with regards to how “strong” the instruments are. Legislative (and economic for that matter) are regarded as sharper than supplying instruments. Information might be considered as the weakest form.

**Policy packages effect**

However, it is vital to have a broad perspective in analyzing policy packages. A long range of literature argues that multi-instrumental instruments have greater potential to reach transport objectives. Thus, a combination of instruments is more likely to perform better than single use of instruments. Consequently, instruments will have different target groups and affect and motive people differently and often differently change the underlying causes of travel behavior (Eriksson 2008). For instance, information campaign urging people to reduce cars can have stronger impact on those with strong environmental beliefs, while economically instruments reducing car-use might influence other segments of the society. Therefore, a combination of instruments is often more effective. Then, it might also be important to analyze when and to which extent do (a combination of) instruments influence travel behavior.

Instruments effectiveness can be evaluated in several ways. They can be effective on their own (and more probably more so if it is a direct and strong instrument), but can also be merely ineffective. An important aspect is to clarify when and why an instrument is ineffective, but also to investigate how a policy package can increase effectiveness. Some instruments can in it selves be ineffective, but combined with other instruments could have strong impacts. Columns at the right of the table illustrate such a point. A major point is to clarify in which policy packages provides greater benefits combined than used separately. May et.al (2006) structure of synergies provides a useful classification on how policy instruments might interact. For simplicity, combination of instruments can be synergic or additive. Positive synergy is present when two instruments provide greater benefits than used separately. This can also be negative if a combination
results in less benefit than using them separately. Additive means that the sum of gains from each instrument is equal to the gain of using them in isolation\(^3\).

Adverse effects are also another important aspect to analyze in a policy package. An instrument (or instruments) can experience unintended effects which can be positive or negative. In combination, two or more instruments may also result in contradictory effects if they are at odds with each other. This can be highly relevant. According to OPTIC (2010), adverse unintended effects often occurred in transport policies.

Instruments can also function as pre-conditional (Givoni et al. 2010). In order to be effective, such instruments rely on implementation of additional instruments. By itself, it will not function or have marginal effects. In a policy package, indirect instruments could have a higher probability of being pre-conditional, since they may induce change by making alternatives more favorable.

Instruments are defined as redundant if they have no or minor effect to a policy package. Thus, such instruments are clearly unnecessary. In addition they may instead represent additional administrative costs (ibid). Pre-conditional instruments might also be classified as redundant if no additional instruments are employed.

<table>
<thead>
<tr>
<th>Table 1 Classification of instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Direct</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Indirect</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Bergen as case**

Bergen is a city located in the western part of Norway and has about 256 000 inhabitants. According to Norwegian pollution regulations, NO\(_2\)-levels shall not exceed 200 micrograms per m\(^3\) more than 18 hours each year. In addition, average level of NO\(_2\) per m\(^3\) shall be below 40 micrograms for the whole year. Figure 1 illustrates the highest registered levels of NO\(_2\) between 20\(^{th}\) of December and 31\(^{th}\) of December.

\(^{3}\) Instruments can be measured in terms of effectiveness in short and long term. In this case, short term effects were vital and instruments will be evaluated accordingly. A policy package can also provide synergic effects in terms of their joint impact on users, more financially feasible and socially or politically acceptable (May and Roberts 1995).
of January. It was registered more than 93 hours of exceeding levels in the period between 8th and 15th of January and over 134 hours for 2010 in total. Thus, pollution levels were well above national regulations. Moreover, Bergen has throughout the last 10 years had an average level of nitrogen dioxide above 40 micrograms. This development is not likely to change. According to the Norwegian Climate and Pollution Agency, bigger cities in Norway will probably exceed national regulations the next 10 years (KLIF 2010).

Figure 1 NO₂-levels between 20.12.2009 – 31.01.2010

There was therefore great pressure against the local government to reduce local air pollution and local authorities felt compelled to intervene. Local newspapers and politicians called for action and suggestion of possible instruments and general critics were raised. However, instruments to reduce levels of nitrogen dioxide were not initiated until 12th of January. A crisis team was assembled, consisting of a broad specter of actors. Police, road authorities, operators of public transport, administration and the county were all represented. In beginning of February levels of nitrogen dioxide was again causing high levels of pollution. Thus, it was necessary for the local authorities to initiate a second policy package.

Course of event

Table 2 illustrates which instruments that were employed, when they were initiated and phased out. In addition, instruments are classified as direct or indirect and whether they are legislative, supply or informative.
Bergen municipality initiated a broad range of instruments. From 12th January direct express busses were established between central nodes. Park- and ride facilities were created in relevant areas. Application was sent to the Norwegian Road Authorities for allowing high-occupancy lanes in all highways into Bergen. The purpose was to increase accessibility for public transport and cars with one or more passengers. In addition, a massive information campaign was initiated with own web pages, full-page and front-page advertisements in local media, SMS to the public, letters to all households and letters to parents in schools and day-care centers. All of these instruments can be classified as indirect. None of them are legislative, but facilitative and informative.

As pollution levels (and traffic level) did not decrease, further instruments were deemed necessary. The following day the police agreed to control idle running, HOV-lanes from north of the city was initiated and parking restrictions (capacity in public parking spaces greatly reduced and public parking spaces in streets were removed). Especially parking restrictions can be regarded as restrictive. Difficulties in parking vehicles (or considerable expensive) can potentially be an important instrument reducing traffic. This is also the first instrument in the package which was directly targeted at reducing traffic. HOV-lanes are classified as an indirect instrument. Its main purpose was not directly aimed at cars, but rather improving accessibility for public transport and cars with one or more passengers. It can also give incentives to car pooling. In this case study, it was important to reduce traffic in short term.
Agreement with the police is characterized as an indirect instrument. Its focus was targeted against inefficient driving, and did not serve as an incentive in order to restrict traffic.

Despite some effort, Bergen still had pollution levels well above national regulations. Therefore, on 14\textsuperscript{th} January 170 000 SMS were sent to the public informing them about odd-and-even number restrictions the following day. These instruments have two significant differences. Although, SMS also were sent on the 12\textsuperscript{th} of January, it was first and foremost designed as a campaign in order to make people not use car. On 14\textsuperscript{th} January information was sent in order to report about a coming instrument. It can be understood as an instrument which was necessary to notify the public about odd-and-even numbers. Thus, they had two different purposes.

The following day odd-and-even numbers were initiated. This instrument is clearly the most restrictive used in the policy-package and aimed directly at car-drivers. Theoretically, despite some exceptions, the number of cars passing into Bergen should be halved. However, it is necessary to point out that the police clearly stated that they would not pursue any controls since they did not have the juridical authority to impose fees. The police could only commence criminal proceedings, and this was regarded as too severe. Odd-and-even numbers were only employed for one day. A couple of days later another HOV-lane were opened and was the last instrument employed in the first policy-package. At 4\textsuperscript{th} February all instruments were phased out.

The second policy-package was initiated on 4\textsuperscript{th} February. The package was, at large, a replica of the first version. There were only three instruments that were not used; odd-and-even numbers, SMS to the public and parking restrictions were less severe. Consequently, the instruments in the second policy-package was less directly aimed at reducing traffic or characterized as legislative.

Overall, the majority of instruments employed were soft, indirect and more aimed at facilitating public transport and campaigning (informing). Parking restrictions and odd-and-even numbers are restrictive measures, but they were to a less extent used. They were also phased out earlier than the other instruments.
Little effects

Figure 2 illustrates the number of cars driving into Bergen between 11th January and 18th February. The traffic survey is used in order to analyze how effective instruments were in reducing the number of cars.

Stability seems to be a major conclusion based on results from figure 2. There is a weak tendency to a slight reduction of passing cars on 12th and 13th of January (5%). This was at the beginning of the first policy-package with an intense information campaign. However, a much larger reduction was necessary in order to significantly reduce the level of nitrogen dioxide.

On 15th of January significantly fewer cars entered Bergen. Between 6 and 9 a.m., a traffic reduction of 27% was achieved compared with the day before and Fridays later in the period. This was by far the most efficient instrument in terms of reducing traffic.

The following days can be characterized as rather stable. Traffic levels are quite steady and on a relative high level the next weeks. Interestingly, there are no major differences in traffic levels between policy-package 1 and 2. If the instruments had a significant effect, we would expect more traffic between these packages. Traffic levels are also highest in week 7, which is during the second policy-package. Therefore, at large, evidence point to the fact that instruments had minor or no effect. The only exception was odd-and-even numbers. Policy-packages aimed at reducing traffic were rather unsuccessful. Stability is the keyword describing effects of the different instruments.

In table 3 instruments are classified and effectiveness is considered. Single effect means to evaluate instrument effectiveness isolated from other instruments, but also their effect in combination is considered.

![Figure 2 Passing cars towards Bergen centre](image)

Figure 2 Passing cars towards Bergen centre

Based on a report from National Public Road Administration (2010)
Table 3 Classification of instruments and effects

<table>
<thead>
<tr>
<th>Instruments employed</th>
<th>Precision</th>
<th>Typology</th>
<th>Single effect</th>
<th>Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct express busses</td>
<td>1</td>
<td>3</td>
<td>( -)</td>
<td>1</td>
</tr>
<tr>
<td>Park-and-ride</td>
<td>1</td>
<td>3</td>
<td>-</td>
<td>x</td>
</tr>
<tr>
<td>HOV-lanes from south of the city</td>
<td>1</td>
<td>3</td>
<td>( +)</td>
<td>x</td>
</tr>
<tr>
<td>Information to the public and companies</td>
<td>1</td>
<td>3</td>
<td>( +)</td>
<td>x</td>
</tr>
<tr>
<td>Agreement with the police controlling idle engine</td>
<td>1</td>
<td>3</td>
<td>-</td>
<td>x</td>
</tr>
<tr>
<td>HOV-lanes from north of the city</td>
<td>1</td>
<td>3</td>
<td>( -)</td>
<td>x</td>
</tr>
<tr>
<td>Public parking spaces are removed</td>
<td>x</td>
<td>x</td>
<td>( -)</td>
<td>x</td>
</tr>
<tr>
<td>Capacity in public parking houses are greatly reduced</td>
<td>x</td>
<td>x</td>
<td>( -)</td>
<td>x</td>
</tr>
<tr>
<td>170 000 SMS to the public informing on the situation</td>
<td>x</td>
<td>x</td>
<td>( -)</td>
<td>x</td>
</tr>
<tr>
<td>Odd-and-even numbers</td>
<td>x</td>
<td>x</td>
<td>( +)</td>
<td>x</td>
</tr>
<tr>
<td>HOV-lanes from the west of the city</td>
<td>1</td>
<td>3</td>
<td>( -)</td>
<td>x</td>
</tr>
</tbody>
</table>

Discussion

These results make it possible to highlight several points and illustrate some patterns. First, few instruments had significantly results on traffic levels in isolation. This means that most of the instruments were dependant on synergic effects in combination with other instruments. Odd-and-even numbers are the only case which clearly demonstrated a travel reduction. This is in line with Marshall and Banister (2000) which demonstrated that, without restrictive instruments, promotion of alternative forms of transport will probably be less effective. The distinction between synergic and additive effects in a policy package is important. A major advantage of a policy package is to foster synergic effects. However, available data makes it rather impossible to analyze whether an instrument had a synergic or additive effect. It is not possible to isolate an instruments effect and analyze it without taking into consideration that it was employed in combination. We do not know what the traffic levels would have been like if, for instance, odd-and-even numbers were initiated as a single instrument. In addition, categories are not necessarily exclusive. Instruments can both be synergic and have adverse effects. In this case study, only limited positive effects were registered. Consequently, synergic effects are also small. Despite these limitations, we cannot exclude the possibility of such mechanisms. Synergic effects could be linked to a combination between direct, restrictive instruments and indirect, weaker instruments. Restrictive instruments, as odd-and-even numbers, could reinforce the effect of, for instance, direct express busses.

Second, the results indicate a pattern of a possible correlation in effectiveness between direct and indirect instruments. Few of the indirect instruments proved to drastically change travel behavior. Although, they had some positive isolated effects, it was not sufficient to alter either pollution or traffic levels. However, not all of the direct instruments were successful either. Reduced parking capacity had
only marginal effect. In this regard, parking restrictions might be an instrument which needs some time to impose change. In the longer run it might, however, have a larger impact. Private companies also own a large share of the total number of parking areas in Bergen. This implies that it was not sufficient to only reduce public parking space. In the end, it is hardly a coincidence that odd-and even numbers were most effective and being classified as a direct and legislative instrument. If we limit our analysis to only single effects, direct and restrictive instruments are probably more effective than indirect and weaker instruments. These results are compatible with O’Fallon et al. (2004). They conducted a stated preference statement in order to look at policies which can make drivers to change their travel behavior. They argue that policy packages will be most effective to constrain car driving. In addition, sticks showed to have greater impact on travel behavior compared to carrots.\(^5\)

Third, adverse effects were registered in the majority of instruments in a wide range of factors. Direct express busses did increase the public transport supply, but they did also contribute to increased pollution. Diesel engines are a major polluter of local air pollution. Therefore it is necessary that busses have a high utilization factor in order to effectively reduce pollution. In the case of Bergen many busses had low utilization, which could lead to increased pollution.

HOV-lanes were established on three main roads into Bergen; from south, west and north. The main result is that HOV-lanes may result in some reduced level of traffic, and cars with passengers had some reduced travel time. This was especially the case for the northbound and southbound lanes. However, some problems with regards to traffic safety were also registered. Both police and road users reported of near accidents due to HOV-lanes. In total, despite increased accessibility due to HOV-lanes, effects on air pollution are uncertain. Cars which didn’t have passengers experienced longer travels which could offset the reduced emissions from vehicles with better accessibility. Such aspects reduce the overall effectiveness of the policy package.

Fourth, there might be a pattern between indirect, weaker instruments and whether they function as a pre-conditional instrument. Establishment of HOV-lanes might be a pre-requisite for strengthening public transport’s accessibility. Strengthening public transport can also be a pre-requisite if local authorities decide to restrict the

\(^5\) However, indirect instruments can be important in order to reduce adverse effects and increase legitimacy, but such aspects are not analyzed in this paper.
use of cars. As a single instrument they are likely to have low effects, but can function as a necessary instrument if the policy package should include stronger and more restrictive instruments. Some redundant instruments were also registered. Usually there are no queues in the main road from west in Bergen and the traffic flow is quite good. However, when the HOV-lane from west was established, travel time increased and queuing was registered. In order to reduce pollution, this instrument resulted in contradictory effects. Agreements with the police were also an instrument which probably was redundant. Idle-running cars are not the main problem in Bergen when it comes to local air pollution, and it gives no incentives to reduce car traffic.

Also, at the political level, it seems to be that politicians tried to extensively use soft and indirect instruments. According to OPTIC, this can be a prudent strategy. To provide the carrots before the sticks can increase legitimacy and may be a necessity in order to provide alternative transportation. Also, politicians can have a strong tendency to successively move from the least restrictive instrument to more restrictive instruments (Bemelmans-Videc and Vedung 1998:264). This was clearly the case in Bergen. As most instruments proved not to be effective, stronger and restrictive measures was deemed necessary. However, the sudden high level of pollution also revealed that municipalities did not possess adequate instruments to effectively reduce traffic. According to Norwegian laws, local authorities do not have the statutory authority to either implement odd- and even numbers or increase charging in the toll rings. In order to implement odd- and even numbers, an exemption had to be made by the national authorities. Thus, it was challenging for local authorities to implement instruments which were direct and restrictive. Such mechanisms would probably increase the overall effectiveness of the policy package.

**Conclusion**

In the end, this case study makes it possible to highlight several important lessons. First, pollution levels struck as a surprise among the local authorities and a specific emergency plan aimed at reducing traffic did not exist prior to the policy package. Consequently, there was not made any preceding analysis of recommended or available instruments which could be implemented in such a situation. Local authorities had to develop a new emergency plan. This had important consequences for the final design of the policy package. Policy makers did not have any apparent or evident understanding of each instruments effect. It is therefore not surprising that several redundant instruments were initiated and
that contradictory effects were registered. The majority of instruments had only effect in combination with instruments that were restrictive and were directly aimed at reducing transport. Without such, they could be classified as ineffective. This means that they had little or no effect beyond that existed before its implementation. Therefore, the case of Bergen illustrates a situation where a broad specter of instruments was employed, but the overall effectiveness was limited. It also reviled that legislative changes is necessary in order to provide municipalities adequate instruments to alter such a situation.

Second, without prior plans, it is probably more likely that politicians initiate a policy package which is sub-optimal in terms of effectiveness. Bergen implemented a policy package which was both too extensive and too indirect in terms of reducing traffic. An extensive package can have important consequences in terms of increased complexity and costs. Fewer instruments would involve fewer actors in the process of designing and implementing a policy package.

Costs is also entailed each instrument. The major focus was to reduce the level of cars, and expenses related to each instrument were of secondary importance. However, an extensive and partly indirect policy package will increase the overall costs. Initiating ineffective instruments can therefore result in including unnecessary new actors and lead to avoidable public expenses. The case study illustrated that several instruments were expensive and rather ineffective in terms of effectiveness.

Givoni et al. (2010) points out that “there is a danger that packages may themselves be tacitly and mistakenly revered (…)” and ““given the limited material, evidence available to their success, policy packages should not be constructed as a panacea to the complexities and challenges of policy making in their own right”. This paper illustrate that a policy package is necessary in order to effectively reach objectives. However, it is vital that policy makers have prior understanding of the interaction between the different instruments. It seems that too little consideration was put towards the potential between instruments and the likely effects. A smaller, targeted and more coherent policy package could have increased effectiveness.

In total, traffic reduction strategies was rather ineffective. Indeed, soft policy instruments can be important in order to make restrictive instruments more socially and politically acceptable. In this way policy makers can overcome barriers in developing effective policy packages. However, it is not possible to analyze such aspects in this paper.
References

Action Plan for increased air quality (Handlingsplan for bedre luftkvalitet) (2007) Bergen municipality
https://www.bergen.kommune.no/bk/multimedia/archive/00098/Handlingsplan_fo_r_be_98555a.pdf


Eriksson, L. (2008) Pro-environmental travel behavior: The importance of attitudinal factors, habits, and transport policy measures. Departments of Psychology, Umeå University, Sweden

Givoni, Moshe, James Macmillen and David Banister (2010) From individual policies to policy packaging Submission to European Transport Conference 2010


National Public Road Administration (2010) Registrering av trafikkforhold. Erfaringer med strakstiltak ved ekstraordinær luftforurensing i Bergen vinteren 2010 (Registration of traffic. Experiences with immediate measures in Bergen winter 2010)

OPTIC (2010) Inventory of measures, typology of non-intentional effects and a framework for policy packaging

http://optic.toi.no/mmarchive_getfile.php?mmfileid=14934&CPMMFILEID_URL_WYSIWYG_TOKEN=1


Sørensen. Claus Hedegaard, Karolina Isaksson and Jonas Åkerman (2011) *Adoption and implementation of policy packages in transport – barriers, success factor and strategies* Paper for the OPTIC workshop at the TEMPO conference, Oslo, February 1,


Dr.-Ing. Dipl.-Wirt. Ing. Stefan Tetzner
Federal Highway Research Institute (Bundesanstalt für Straßenwesen)
Department V2 “Verkehrsstatistik, BISSta“ (Referat V2 „Verkehrsstatistik, BISSta“)
tetzner@bast.de
+49 2204 43-526
Brüderstraße 53
51427 Bergisch-Gladbach

Achslastdatenerfassung auf deutschen Autobahnen
Weigh-in-motion on German motorways (final paper)

Topics:

Abstract 2

1 Axle load data collection 3
1.1 Background 3
1.2 Axle load data collection in Germany 4
1.3 Procedure of axle load data collection 5
1.4 Collected data 5
1.5 Measurement problems 7

2 Results of the axle load data evaluation 8
2.1 Overview 8
2.2 Overloading of Vehicles 9
2.3 Relations between DTV_{SV} and Load Volume 10
2.4 Direction-related axle load distribution 11

3 Prospects 13

References 15
Abstract

In Germany, traffic volumes are permanently registered at over 1,300 stationary automatic counting stations. These are supplemented by manual traffic counts on almost all road sections of the federal road network every five year. It is less well known that also the axle loads of the vehicles on the federal motorway network are recorded at 21 stationary axle load weighing stations. For the future, it is intended to add mobile axle load scales. These are planned to be employed for limited time periods at bridge structures of federal motorways as well as on federal trunk and rural roads. The accuracy and the feasibility of such a system is currently tested in a test field.

The main evaluation data of axle load measurements are total vehicle weights, axle and axle group loads, vehicle types and vehicle speeds. Further weighing station specific parameters can be derived from these data, such as vehicle distribution, equivalent 10t axle passes as well as load collective quotient $q_{Bm}$ that can be considered as a characteristic value for the section-specific road stress. Main objectives of axle load data collection are the forecast of the temporal and spatial development of the loads acting on the federal trunk road network, an economic and reliable maintenance planning for the road and bridge structures, improved approaches for dimensioning the road pavements and bridges, as well as information about the distribution of vehicle types, axle loads, total weights and overloads in the federal trunk road network.

Based on the major results of the past axle load measurements, the following conclusions for German motorways can be drawn: 1) increasing number of overloaded vehicles or overloaded axles; 2) evidence of a linear relation between the traffic volume and the load or the 10t equivalent within a direction-related axle load weighing system; 3) in some cases clear differences in the vehicle weight or the axle load distributions between the two traffic directions at an axle load weighing system.
1 Axle load data collection

1.1 Background

In addition to the traffic volumes registered on German federal motorways and federal roads by over 1,300 automatic counting stations and manual road traffic counters used in five-year intervals, the axle loads of heavy-duty vehicles are determined at some cross sections of the federal motorway network. In the future, the collective of currently 21 stationary cross-sectional axle load weighing stations (see Figure 1) will possibly be extended by mobile axle load scales. They can be employed for limited time periods at bridge structures and federal motorways as well as on federal trunk roads and rural roads.

Figure 1: Network of axle load weighing stations in Germany
The main objectives of axle load data collection are

- improved efficiency in the construction and maintenance of roads, bridges and engineering structures,
- increased traffic safety,
- further development of the efficiency in road use,
- pre-selection of overloaded vehicles.

The collected data are used for the further development of the National German Guidelines for Road Layer Thickness Measurement (Guidelines for the Standardisation of Superstructure of Road Surfaces; in short: RStO), and they represent the basis of idealised load models for analyses of the stability and fatigue strength of bridges and engineering structures as well as various statistical evaluations.

### 1.2 Axle load data collection in Germany

The objective is to establish a network of 40 stationary cross-sectional axle load weighing stations on German motorways. At the weighing stations, the number of passing vehicles are counted and the loads of the passing axles as well as the gross vehicle weights are determined usually at the main lanes in both directions – in some cases also on the first fast lane – and the registered data are saved.

In the first expansion stage, five units of bending plate design were installed. For the subsequent nine weighing stations of the second and the seven axle load scales of the third expansion stage, the piezo-quarz technology will be employed. Due to meanwhile outdated and unreliable weighing technology and obsolete electronic systems, the axle load scales of the first stages are to be replaced by more advanced systems based on Piezo technology.

In the electronic system of the first-generation weighing stations, removable storage media with a very low storage capacity, compared to advanced units, were used. These devices had to be exchanged regularly by blank data carriers. Due to the restricted storage space, a compromise between an acceptable change cycle and the saved data contents had to be found. For this reason, in the initial stage of axle load measurement, the data were collected for an hour before they were actually saved.

Thanks to advanced electronics, larger storage media and the possibility to transfer data online, the latest load weighing stations are capable of indicating the data of individual vehicles. If possible, a daily data transfer through GPRS (General Packed Radio Service) or UMTS (Universal Mobile Telecommunications System) to BASt is intended in the future. With the latest units, cars and special vehicles are detected in addition to standard trucks.
Special vehicles are often characterised by extreme total weights. The resulting data pool is of particular interest for the dimensioning of bridge structures and road pavements as well as for maintenance management.

1.3 Procedure of axle load data collection

The axle load strip sensors currently installed in Germany basically consist of the following three components:

1) In the latest models, a piezo-quarz element is used as sensor. Piezo-electronic sensors consist of quartz crystals, which generate and increase electric voltage in proportion to the acting load without causing deformations. The sensor is grouted into a sand epoxy compound in a milled slot in the roadway, vertically to the traffic direction (Figure 2).

2) Directly before and after the axle load sensor, an induction loop is arranged in the roadway. This will be used to detect the speed and type of the passing vehicle.

3) Outside the roadway, the electronic systems for the sensors are arranged on one side of the detection cross section. Depending on the number of sensors, up to 12 lanes can usually be covered by the detection facility. The electronic system is responsible for the processing, storage and transfer of the data. A stationary power supply ensures that the system operates without any interruptions.

Figure 2: Structure (left) and installation (right) of a piezo-quarz sensor (Kistler, 2011)

1.4 Collected data

By combining inductive loop detectors with axle load sensors, it is possible to distinguish heavy-duty vehicles. With the two sensor technologies, the measured axle loads can be assigned to the vehicles so that the following information is available:
• all individual axle loads of a vehicle,
• axle spacing to each other,
• speed of a vehicle,
• vehicle length,
• distance between vehicle and vehicle travelling ahead,
• time of measurement.

Based on the directly recorded data mentioned above, the electronic system calculates further data that will be included in the dataset:

• total weight of the vehicle,
• possible total overload (according to § 34 Road Traffic Registration Ordinance; in short: StVZO)
• type of the axle assembly (single, double or triple axle),
• possible overloading of single axles or axle assemblies as well as
• vehicle type (according to the guideline “Technical Delivery Conditions for Route Stations 2002“; in short: TLS 2002)

To ensure that usable results are obtained for subsequent technical evaluations, the measuring accuracy must be within defined tolerances. To this end, the weighing stations need to be maintained and calibrated regularly. To check the measuring accuracy of axle load weighing, statically pre-weighed trucks are used which pass the weighing station several times. Calibration of the weighing station is likewise required directly after sensor installation prior to system start-up. The calibration and the test plan is applied according to the COST 323 definitions. An accuracy of class B+(7), at least B(10), has to be achieved.

The data transferred to the individual federal states or the Federal Highway Research Institute (BASl) are statistically evaluated and processed accordingly. For this purpose, it is necessary to derive the required parameters from the collected data, which are then used to determine the influences on the infrastructure or which may be integrated in new guidelines and road models:

• Vehicle type distribution
• Total / axle weight distributions
• Number of axle passes \( f_A \)
• Equivalent 10t axle pass
• Load collective quotient \( q_{Bm} \) (see Section 2.2 for explanations)

The data acquired with the axle load measurements have various purposes: forecast of the temporal and spatial development in the stress acting on the federal trunk road network, support of an economic and reliable maintenance planning of the road and bridge structures, im-
proved approaches for dimensioning the road pavements and bridges as well as statements about the distribution of vehicle types, axle loads, total weights and overloads in the federal trunk road network. In addition, some axle load weighing stations serve as pre-weighing stations for the roadside checks where vehicles, which exhibit individual overloaded axles or which are incorrectly loaded or overloaded as a whole (see Fig. 3), are directed to the check site and checked. The aim of these checks, where vehicle drivers and owners of not properly loaded vehicles can be sanctioned and forced to reload, if necessary, is to limit damage to the traffic infrastructure, to achieve a positive influence on the traffic accident record and to ensure equal conditions for competition.

*Figure 3: Truck check (left) and incorrectly loaded truck (right; both pictures: BMVBS, 2011)*

**1.5 Limits of the measurement**

The axial load weighing system determines the actual, static axle loads and total weight of the vehicles (whereas the measured data contain the dynamic force by the axles of the vehicles). Since the measurement is conducted at normal vehicle speed with the aid of a sensor integrated in the roadway, vehicle vibrations caused by driving dynamics and hence also vibrations of the weight forces acting on the roadway must be kept as low as possible. For this reason, a high measuring accuracy does not only rely on the sensor, but also on the roadway characteristics; and these requirements must be met even with high traffic volumes. Prerequisite for the installation of the axle load scales in the roadway is, for example, to take care that the road pavement has optimum characteristics. On a section of 300 m before and 25 m after the axle load weighing station, a stable, flat and defect-free surface must be ensured.

In addition to the road condition, a large number of other influencing factors are crucial for the accuracy of the measured axle loads:

Dynamic additional loads are caused by roadway unevenness (build-up of oscillations or vibrations of the vehicle), but also by decelerating and accelerating vehicles. In addition to ensuring a minimum longitudinal slope and evenness of the roadway, cross sections are to be
selected for axle load weighing stations, where few traffic disruptions (traffic jams) are expected.

It was shown in the past that regular calibrations – for example in intervals of six months – lead to a considerably higher accuracy of the measurement results. For calibration, statically pre-weighed vehicles loaded with various weights are driven at a constant speed over the weighing station and the electronic system is adjusted accordingly.

Problems for single or sequential datasets are caused by cross drivers, i.e. vehicles that, for example, pass parts of the axle load scales of two lanes at the same time during a lane change.

### 2 Results of the axle load data evaluation

#### 2.1 Overview

At the Federal Highway Research Institute, a large number of statistical evaluations of traffic data are conducted. The axle load data provide a basis for proper maintenance and construction planning of roads by supplying realistic load data of heavy goods traffic. In addition, they represent a major factor for the load calculations and dimensioning of bridges and enable a sustainable preservation and maintenance of bridge structures.

In the sections below, the focus will be exclusively on the statistical development and distribution of the axle load data. The detailed representation of the above conclusions from the research fields “Dimensioning and Maintenance of Roads“ as well as “Building Maintenance“ will not be considered in this context. Generally with the collected axle load data the B(10) specification according to COST 323 is achieved, but it is strived to reach the B+(7) standard.

The three following conclusions can be drawn, reflecting the main results of the past axle load measurements (sections 2.2-2.4):

1) The historic development shows an increase in overloaded vehicles or overloaded axles. Overloaded axles were particularly found in the two-axle tractor unit with three-axle trailer (Vehicle Type 98 according to TLS 2002);

2) A linear relation between traffic volume and loads, i.e. the 10t equivalent within one direction-related axle load weighing station, was observed;

3) Regarding the vehicle weight or the axle load distributions, clear differences were found in some cases between the two traffic directions at one axle weighing station.
2.2 Overloading of Vehicles

The development of the heavy goods traffic volume as well as the tonnages of the individual heavy-duty vehicles are of special interest for the statement if the road dimensions are still in line with the traffic volume, i.e. the load volume acting on the road. While the heavy goods traffic volume on the German federal motorways increased e.g. between 1998 and 2008 from 28.9 to 34.6 [billions of heavy-duty vehicles - km] (Fitschen, 2010), the loading situation of vehicles likewise shows an upward trend.

For example, the percentage of the vehicle type 98 in the heavy goods traffic volume on federal motorways grew from 35.1 % in 1998 to 49.6 % in 2005-2008. As shown in Fig. 5, an overloading of the permitted total weight of 40 t was found on approx. 20 % of the tractor units of type 98. In 2005-2008, total weight overloads were found even on one quarter of Type 98 vehicles at some weighing stations. At the same time, the percentage of unloaded or only lightly loaded vehicles (in the range of 12 to 19 t) decreased.

Even if Fig. 4 only represents various weighing stations for the two reference years of 1998 and 2004/2005 (no comparison values are available due to faults in weighing stations), the trend towards higher loads on trucks is clearly demonstrated.

Figure 4: Total weight distributions of tractor units (Type 98) at various axle load weighing stations in 1998 and 2004/2005 (Kaschner, 2009)

The axle overloads, which were found, mainly occurred on the driving axle (2nd axle). In the case of the five-axle tractor unit (Type 98), the second axle is overloaded in nine out of ten affected vehicles. This is due to the fact that the cargo on the trailer is usually moved towards...
the front wall. In many cases, no vehicle overloading is present, but only an incorrect load
distribution within the trailer. This axle overloading is hardly detrimental to bridges; however,
it causes significant damage to roads.
This overloading is favoured by the fact that in Germany usually the five-axle Type 98 tractor
unit is used instead of the six-axle vehicle type 106, which is equipped with a double driving
axle.

2.3 Relations between DTV_{SV} and Load Volume

In the period from 2004 to 2008, a correlation between the heavy goods traffic volume (Average
Daily Traffic Volume of Heavy Goods Traffic; in short: DTV_{SV}), measured in heavy
goods vehicles per day and the equivalent 10t axle passes, in short: 10t axle passes, was de-
termined for the available load data from the stationary axle load weighing stations in Ger-
many.

The 10t equivalent is an auxiliary parameter that reflects the actual road stress by exponen-
tially weighting heavy-duty vehicles. It is calculated as sum of the loads per axle load class –
standardised for a 10t axle pass – multiplied with the number of axle passes:

\[
EDTA_{i-1}^{(SV)} = \sum_{k} [DTV_{i-1}^{(SV)} \cdot \left( \frac{L_{\xi}}{L_{k}} \right)^{2}] 
\]

with

\[
DTV_{i-1}^{(SV)} = DTV_{i-1}^{(SV)} \cdot f_{a_{i-1}}^{(SV)}
\]

(RStO 01)

\[DTV_{i-1}^{(SV)}: \text{Average daily traffic volume of heavy goods traffic in the year of use } i-1 \text{ [vehicles / 24h]}\]

\[EDTA_{i-1}^{(SV)}: \text{Average number of the daily equivalent axle passes of heavy goods traffic in the year } i-1 \text{ [vehicles / 24h]}\]

\[f_{a_{i-1}}^{(SV)}: \text{Average number of axles per vehicle of heavy goods traffic (axle number factor) in the year } i-1 \text{ [axles / vehicle]}\]

\[L_{k}: \text{Mean axle load in load class } k\]

\[L_{\xi}: \text{Reference axle load: } 10 \text{ t}\]
As a result of this examination, it can be stated that for the data clouds of separately considered weighing stations, a significant linear relation (the correlation coefficient $r^2$ ranges between 0.9224 and 0.9967) was found (see Fig. 5), although the weighting of the heavy-duty vehicles is exponential. Based on this, the 10t equivalent and hence the road-specific load can be determined solely from the daily heavy goods traffic volume, even if the axle load data are not available, if the typical load or loading level of the vehicles is known – indicated in the diagram by the slope of the point cloud typical for the axle load weighing station. Whether or not this relation applies to widely separate years, remains to be verified by further evaluations of the most recent axle load data.

Figure 5: Correlation between heavy goods traffic volume and 10t equivalent (BASt, 2008)

2.4 Direction-related axle load distribution

As a further result of the axle load data evaluation, it can be shown that at some weighing cross sections some consistent vehicle loads seem to be present in both traffic directions, whereas at other cross sections the specific vehicle loads differ greatly for the two traffic directions.

This statement can be derived from the comparison of the 10t equivalent with the daily heavy goods traffic volume. Fig. 6 and 7 illustrate the correlations between the two values in the period 2004 to 2008 for two weighing stations.

The first example (Fig. 6) shows a weighing cross section at the motorway A 2 (west of Berlin) where the 10t equivalent of the outward and return direction – with an identical daily
heavy goods traffic volume – clearly differ from each other over the whole traffic area. A linear regression over the data determined for the direction of eastward traffic (direction 1) leads to a clearly steeper straight line \( m = 1.43 \) than for the data for the direction of westward traffic (direction 2). The correlation coefficients exhibit a strong relation of the linear regression; with 0.9830 (dir. 1) or 0.9928 (dir. 2), they are situated in the upper range of all weighing stations.

*Figure 6: Heterogeneous distribution of vehicle loads between the two traffic directions of one weighing station*

![Graph](image.png)

Fig. 7 illustrates an almost balanced result in both directions with the example of the data measured at the weighing station “Bahrer Weiher” (motorway A 9 between Munich and Ingolstadt). The two regression lines through the two point clouds exhibit similar gradients with 1.28 (dir. 1/North) and 1.31 (dir. 2/South). The correlation coefficients of 0.9224 (dir. 1) and 0.9571 (dir. 2), however, are in the lower range of the r² values of all weighing stations.

As shown by the two examples, the dependence between 10t equivalent and heavy goods traffic volume of an axle load weighing station is not automatically equal for the two traffic directions. In fact, at some weighing stations the vehicles seem to be heavier or carrying a heavier load in a certain direction compared to the opposite direction. A clear relation to the location, such as a weighing station near a border, cannot be established; however, the number of evaluated stations is too low to provide significant evidence. As a result, the example demonstrates clearly that an extrapolation of the road stress from the heavy goods traffic volumes
only is not possible, without knowing the “specific load” (here the gradient of the regression line between the two parameters) of the – direction-related (!) – cross section. In order to determine this parameter, it is not necessarily required to install a stationary axle load weighing station; in view of the determined continuity of the specific loads of short to medium periods, a time-limited measurement employing a mobile scale is possibly sufficient.

**Figure 7: Homogeneous distribution of vehicle loads between the two traffic directions at one weighing station**

![Graph showing equivalent axle passes as a function of the number of heavy goods vehicles per day. Direction 1 with slope m = 1.31 and Direction 2 with slope m = 1.28.]

### 3 Prospects

For the future, a detailed statement about the stress of the whole federal trunk road network caused by trucks is necessary. In order to provide a precise representation of the load volumes for a specific route, the extension of the axle load weighing station network to federal roads is necessary. For this purpose, the temporarily application of a mobile Bridge-WIM (B-WIM) will be examined which may be used at further motorway cross sections, but also and in particular at federal secondary roads. With the aid of mobile systems which are attached for a limited time period to the lower side of road bridges, axle load data can be measured at a far bigger number of cross sections than this would be possible with stationary axle load scales. The collected data would represent an important basis for the stress-oriented dimensioning of
bridge structures and road pavements as well as for an advanced maintenance management of roads for the complete – especially for the subordinated – road network.

Before Bridge-WIM-systems are applied to a greater extent, the setup of a test field is intended. For the collective it is necessary to choose only one-span beam bridges with a range between 15 and 25 meters. The tests are expected to provide reliable information, such as

- accuracy of the achievable axle load data,
- time consumption for the first installation and regular installation and dismantling of the mobile B-WIM-system,
- procedure of installation and operation with flowing traffic,
- service life of the system,
- durations of use based on the determined axle load data – particularly on federal roads.

On the basis of the results, the cross sections where the systems are to be placed and the operation intervals and periods of the mobile Bridge-WIMs are to be defined. The further extension of the stationary axle load weighing station network should be planned depending on the results.

With the aid of time-limited axle load measurements, direction-related load levels of heavy-duty vehicles can be determined (see Section 2.2). According to traffic volume counts, as they are performed at over 600 cross sections on federal roads with the aid of automatic counting stations, and on the basis of the manual road traffic measurements, it will be possible in the future to perform extrapolations of the 10t equivalent for the relevant road sections.
References


Integral consideration of the lightweight design for railway vehicles
1 INTRODUCTION

Optimizing the efficiency of all transport vehicles requires comprehensive and systematic solutions, particularly when the objective is to reduce energy consumption connected with CO₂-emission. Driving resistance and reducing mass play key roles in this effort. The lightweight design of railway vehicles leads to numerous primary as well as secondary advantages. The primary effects are e.g. energy reduction and adherence of the maximum load per wheel set. Secondary effects include reduction of wheel sets itself what leads to a weight reduction of the trains and the aerodynamic resistance.

A comparison of road and rail vehicles reveals that rail vehicles must move more mass per passenger than road vehicles. Reasons for this include high requirements on static and dynamic loads (DIN EN 12663, DIN EN 15227) and a configuration of the vehicles for greater mileage. Car body weight and the required drive power are mutually influencing. If operational requirements (train schedule) demand a specific acceleration capability, then the required drive power will increase proportionally to vehicle mass. This implies that a secondary effect of a lower vehicle weight is a reduction of the installed drive power, resulting in additional energy savings.

The first section of this paper discusses opportunities for reducing the mass of rail vehicles. We also uncover additional benefits and additional effects (besides reducing energy consumption) of lightweight construction. Using this as a foundation, we introduce in the second section a methodology which describes how the car body of railway vehicles can be optimized in terms of mass. This methodology is developed by DLR within the framework of the "Next Generation Train" project (Figure 1).

How showed in (Winter and Granzeier, 2011) in the context of the project “Next Generation Train” (NGT) different scientific issues are considered. The areas are aerodynamics, structural dynamics, dynamic performance, propulsion technology, material sciences and lightweight construction. The maximal operating speed is 400 km/h with the same safety standards at least. Another aspect is the reduction of the specific energy (energy per seat) at of 50 percent.

![Figure 1: Design image of next generation train](image)

In the last section the concept of a novel load bearing car body structure is defined. The structure bases on the result of the topology optimization and promises a significant weight reduction potential.
2 OPPORTUNITIES OF MASS REDUCTION

The weight reduction of railway vehicles has different impacts. One result is the energy saving. Simulations have shown that the potential for savings through reduction of vehicle mass is dominated by the characteristics of the service profile. There is a particularly high potential for savings with service profiles that have short distances between the stations and low maximum speeds. As the maximum speed increases, the proportion of energy needed to overcome aerodynamic resistance increases and the potential for savings goes down. The potential for savings with Diesel powered railway vehicles (Diesel Multiple Units (DMU)) is up to four times greater than with electrically-powered vehicles because electrically-powered vehicles can recover a portion of the kinetic energy during braking. However, if the electric braking power is not adequate to achieve the required deceleration, then the potential for savings increases with the electrically-powered vehicles as well (cf. Dittus et al., 2011).

Beside the energy saving because of a reduced vehicle mass the maximal axle loads are defined (cf. EBO § 10 (N.N. 1967)). The reason for this is to limit the maximal static and dynamic forces on the rail tracks. The axle loads are a result of the relation of the vehicle total weight to the number of wheel sets. Regarding the static loads the limits depend on the operation purpose and the maximal speed. According to TSI rolling stock, trans-European high-speed rail system (N.N. 2008b) the static axle load is between 22.5 tons for normal speeds and 17 tons for speeds above 250 km/h. The compliance with these requirements is often a challenge. During the run of a train there are also dynamic loads which have an effect on the rail position error and the position of the wheels relative to the rails.

Weight optimised vehicles with unchanged vehicle properties, like the vehicle dynamics, leads to a reduction of the rail damage because of fewer impacts. Lighter axle loads also leads to a reduction of the abrasion regarding wheels and rail heads. Altogether this is a positive aspect on the Life Cycle Costs of the Infrastructure (Rochard and Schmid, 2004).

All over the world railway lines are built for high and very high speed trains only. Sometimes for locomotive-hauled trains it is not possible to run there, that is why on these lines operate only train sets. If it would be possible to reduce the weight of all the trains running on these lines, the superstructure and e.g. bridges could be adapted on the lighter axle loads what leads to an economical benefit.

Because of the individual originated railway lines in the urban traffic (e.g. metros) the axle loads are often fewer than for standard railway lines. A large part of the ground area of the vehicles is purposed for standing passengers. According to DIN EN 12663 (N.N. 2000) a load of 5-10 passengers/m² is expected. This leads to a challenge in respect to the axle load limit and a light weight construction becomes very important.

Lighter car bodies, and consequently lighter vehicles, permit the use of less powerful and lighter drive and braking equipment. In turn, this leads to the use of lighter rotating masses, such as the rotor in the drive motor or the braking discs. The braking discs, which on high-speed trains can weigh more than 100 kg each, are arranged in multiple assemblies directly...
and unsprung on the wheelset shaft. Furthermore, the force of the car body is absorbed by the running gear and directed through the wheelsets and wheels into the rails. Therefore, when the car body mass is reduced, the running gear can be adapted to the car body load and made lighter. This leads to a favourable condition for the dynamic behaviour of the vehicle. So it is possible to reduce the impacts on the track and the costs of maintenance can be reduced.

Despite the active safety measures common to railways, the possibility of collisions between railway vehicles or with obstacles cannot be completely excluded. The resulting energy must be absorbed by crash elements in the vehicles according to DIN EN 15227 (N.N. 2008a). Therefore, in the event of a collision, the mass of the colliding vehicles plays a crucial role in addition to the speed of those vehicles. How showed in (Rochard and Schmid, 2004) this means that the energy that must be absorbed will be lower with lighter vehicles, so the crash energy absorbers can be designed lighter.

Beside the described advantages the train concept can be modified because of the weight reduction. This gives the chance to reduce energy. A main target for a complete regarded light weight train is the reduction of the number of running gears and axles as far as possible. If the axle load and the number of axles are invariant, the length of the cars and the area for the passengers can be raised with light weight carriages in relation to heavier carriages. It is to consider that the length and the corresponding width of the car body are bordered because of the structure gauge. The possible increase of the car body length can lead to the saving of one car at least regarding the whole train with an unchanging length. This has the impact that running gears and connections between carriages can be reduced. The result is a weight reduction of the train, the optimised area utilisation and reduction of the aerodynamic resistance corresponding with energy saving. The aerodynamic resistance is the main factor for the energy consumption of high speed and very high speed trains (Table 1). Another fraction of the driving resistances is the resistance to rolling. It can be divided in a resistance which depends on mass (e.g. friction of the bearings) and one which depends on mass and velocity (e.g. dynamic resistance because of vehicle oscillations).

![Diagram of driving resistance related to velocity](image)

Table 1: Driving resistance related to velocity
The demand of the passengers for comfort and design increases permanently. The costumer request must be responded for attractive railway transportation. Concrete this may be the interior, thermal and noise insulation, air conditioning, comfort and width of the seats infotainment, etc.. With the pretended requirements a weight increase accompanies and a consequently light weight car body structure is absolutely necessary because of the fixed axle load.

Hence the manifold and extensive effect resulting of the weight reduction of railway vehicles, light weight constructions are a main topic for future trains.

3 CORRELATIONS OF LIGHT WEIGHT RAILWAY VEHICLES

3.1 Mass Distribution of a Railway Carriage

A railway carriage can be divided into three main assemblies (Figure 2): the running gears, car body structure and equipment (like propulsion, interior, car body completion, etc.).

![Figure 2: Main assemblies of a railway carriage](image)

The equipment and car body completion is a main cause of the carriage weight (Figure 3). It is composed of a large number of diverse subassemblies. These subassemblies are predominantly independent from each other, so they have to be regarded separately. Based on that, the global weight saving potential resulting of the several equipments is limited.
Running gears must perform diverse requirements, e.g. stability of the run, fatigue strength, low wear in the wheel/rail contact and moveable parts. So weight saving of running gears is difficult because of the manifold and security relevant requirements. Currently there are several ongoing research activities, e.g. by the DLR (Kurzeck and Valente, 2011), to improve run stability, wear and minimize weight of the running gears. It is to note that beside the maximal velocity the car body weight, equipment weight and payload are significant influence parameters on the running gears. For this reason the weight optimisation of the car body is an important criterion for a light weight carriage.

3.2 Influence Parameters on Car Body Structure

Analyses showed that the mass of the rough car body normally accounts for 15% to 30% of the vehicle's empty weight. Nonetheless the car body structure is in an interaction with different parameters (Figure 4). This means the car body structure has a direct influence on the weight saving potential of these parameters.
The operation purpose defines the payload (because of the requested passenger capacity), the number of doors, the axle loads, the velocity connected with the air pressure during train encounters which the car body has to resist, etc.

The car body structure is also influenced by the car body concepts which can constrain the conception of the structure design. For example conventional car body structures constrain the scope of design. On the other hand a novel car body concept gives the possibility to create an ideal structure for the respective requirements. Another influence factor is the train concept which defines the geometry of the car body in principle, the position and distances between supports, etc.. Furthermore the train concept is in an interaction with the running gears. As noted above the running gear concepts define the maximal weight of the completed car body and so its length. Also the running gear concept has an influence on the design of the connection to the car body structure.

The joining technology goes together with the materials. The utilized materials have to correspond with the joining technology. The lightweight approach of the multi material design (adapted material in adapted positions) brings a large number of challenges regarding the connection of materials with different characters. Materials and joining technologies have an extensive impact on the costs, design, concept and weight of the car body structure.

It is obviously that the car body structure is in interplay with different influence parameters. Therefore, optimizing the mass of the car body is particularly important.

### 3.3 Light Weight Principles

Currently there are different car body concepts in use. These are car bodies in the differential style, integral style and rarely in the hybrid style (Figure 5). The differential style is a metallic framework planked with blank sheets in principle. The integral style uses aluminium extruded profiles which are welded together in longitudinal direction and possibly local reinforced. The hybrid style is a mix of different materials which uses the potential of every material.

![Figure 5: Car body concepts (from left to right): differential style (Bönisch et al., 1995), integral style (Elsner and Schnaas, 1993), hybrid style (Müller, 1999)](image-url)
It is to note, that for actual current car body styles (integral style, differential style) a lot of expert knowledge exists. Therefore further weight reductions are limited regarding conventional constructions. A global use of light weight principles is indispensable. The light weight principles can be divided in:

- material light weight construction,
- function and system light weight construction,
- shape and form light weight constructions.

The notice of these principles gives the possibilities of a significant weight reduction. The highest potential for extensive light weight constructions promises shape and form light weight constructions. In this case the load bearing car body structure is adapted regarding the main load paths. So unnecessary material can be saved and the load bearing structure adapted to the inertial forces in the car body. A methodology created by the DLR is predestined for the design of load adapted light weight car body conceptions.

4 METHODOLOGY FOR LOAD ADAPTED LIGHT WEIGHT CAR BODY STRUCTURES

Car body designs must take into consideration a variety of static loads as well as crash loads. These are defined in standards and directives such as DIN EN 12663 (N.N. 2000), EN 15227 (N.N. 2008a), and UIC 566 (N.N. 1990). The car body is thereby divided into two zones with differing tasks and requirements. As defined in EN 15227 the end zones are designed as crash zones to absorb energy. In crash scenarios specified in the standard, the structure in the middle zone (between the crash segments) will not endure plastic deformation (Figure 6). As a result, an adequate level of protection can be provided for people in the vehicle and damage to the car body localized. However, crash elements must be designed so that the maximum forces occurring between the middle segment and the crash zones during a collision do not exceed the assumed static forces for the middle zone. Because of this requirement, only the static loads are relevant for the design of the middle zone's bearing structure. As a result, this entire zone is suitable for the principle of topology optimization. This provides the opportunity to use software to adapt the bearing structure's geometry to the flow of forces. The resulting structure reveals the main load paths that will ideally form under the specified conditions (material dimensions, etc.) and target parameters (stiffness, stress, etc.).
The topology optimization program is based on FEM with a special algorithm. In principle, topology optimization requires definition of the design space available for optimization. This is derived from the selected external dimensions as well as the geometries of the car body, external design, and necessary internal space. The designer can also define in advance the positions where cut-outs will be required for windows, for example, as well as positions where no material may be removed during the optimization process (Figure 7).

In order to give the topology optimization process adequate leeway for optimization, it is advisable to make the walls of the car body as thick as possible. For the calculation process, defined static forces according to (N.N. 2000), for instance are applied to the car body's basic body and, where appropriate, overlayed. During the process, areas of the car body's enclosure that are subject to no or little stress are weakened during iteration steps by way of localized material thickness reductions. The process continues until the previously defined termination
criteria are achieved. This could be, for example, the maximum permissible stress or deflection. When the optimization is depicted, only those car body zones above a certain material thickness concentration will be displayed (Figure 8). The result is a seemingly bionic framework whose structure is pronounced only in the zones that are needed for fulfillment of the requirements.

![Figure 8: Topology-optimized car body](image)

The resulting framework structure is then subjected to a FEM analysis in order to validate the optimization results. This proves that the structure satisfies the static strength and stiffness requirements. An interpretation based on the resulting stress distribution examines the relevance of the sub-zones of the created framework. From this is inferred which of the zones play a role in the overall context and which can be eliminated.

The resulting framework provides information on how an ideal structure should appear under the set conditions and target parameters. From this, we can draw conclusions for the design of the car body concept regarding where bearing structures, surface elements, or joints should be included. Nevertheless, the actual design implementation must reflect the practicality of things like suitable structural components, manufacturing capabilities, cost-effectiveness, and joining technologies.
5 Novel Load Bearing Car Body Structures

Based on the solutions of the topology optimization as well as basic conditions and requirements the Institute of Vehicle Concepts develops novel lightweight structures and car body concepts. It was financially supported by Bombardier Transportation, Division Passenger. The car body must resist different load cases in vertical and longitudinal direction as well as forces resulting of torsion and bending loads. Considering the topology optimization a car body structure was created. It is a self supporting bending and torsion proofed comb tube (Figure 10). The car body structure is made of 3-D cranked bulkheads and sole bars which are trough going. The cranked bulkheads consist of vertical and horizontal beams as well as diagonal beams. The sole bars are proper to mount the longitudinal forces. The vertical and horizontal beams bear the vertical forces. They and the sole bars are connected with diagonal beams. These beams carry the loads and split them in the adjoining beams and sole bars. The diagonal beams are positioned in the window areas of the two carriage levels. Therefore they are predestinated to transmit shear stress and the longitudinal and vertical forces. Against the requirements the beams and sole bars can be made of aluminum, steel or pultruded fiber reinforced plastics. The connection of steel and aluminum beams and sole bars can be made by welding. Another possibility is the use of separate
designed node connections. The respective beams can be suited regarding their stiffness, wall thickness and geometry. According to the result of the topology-optimization the performances of the beams of the bulkheads can be adapted regarding the local internal loads. The principal outline is identical of the beams and the bulkheads along the car body. Therefore a scalable car body in longitudinal direction can be realized.

![Car body structure](image)

*Figure 10: Weight optimized and self supporting bending and torsion proofed car body structure*

### 6 SUMMARY

The reduction of vehicle mass has primary effects on energy saving and reduced axle loads. In addition there are secondary effects like the reduction of the rail damage or an adaption of the superstructure according to the axle loads. Furthermore the train concept can be modified and proved regarding the requirements as a result of weight saving.

The carriage can be divided in main assemblies in which the car body structure takes a key role. It influences manifold parameters and subassemblies. Based on these, lightweight construction of the car body is particularly important. One method of reducing the weight of car bodies (through the use of topology optimization) was described, whereby the main load paths for the static-designed car body zone were determined. The resulting framework structure points towards an economical and weight-saving design while reflecting an optimized flow of forces.

Considering the results of the topology-optimisation a novel car body structure was conceived. It is a self supporting bending and torsion proofed comb tube with significant improvement of the performance.
REFERENCES


Early increase of noise from newly laid road surfaces

Jens Oddershede
Researcher, Danish Road Institute/Danish Road Directorate, Guldalderen 12, 2640 Hedehusene, Denmark, jod@vd.dk; www.roadinstitute.dk, phone: +45 7244 7000

Summary
In order to facilitate the introduction of noise reducing wearing courses, the Danish Road Directorate has developed a system for noise classification of road wearing course products; the so called “SRS system” (Noise reducing wearing courses). The SRS system classifies the noise reduction of newly laid wearing courses into classes compared to a reference wearing course. The noise is normally measured by the Close-proximity method [ISO/CD 11819-2]. From literature there is no knowledge on how the noise emission of newly laid wearing courses changes over the first months. There have never been any guidelines to how old the new wearing course should be after paving before measuring the noise. In this project, newly laid wearing courses are measured weekly with the CPX method within the first months of their lifetime to obtain knowledge about how the noise changes and when the noise characteristics become stable. The project focuses on the noise trend of the 5 new SRS wearing courses and a new reference wearing course with 6 CPX measurements performed over a period of two months after the construction of the wearing courses in August 2010.

The first analyses show a noise incensement in the first months and then it becomes stable. On the background of these results, a recommendation for when to perform noise testing of wearing course products in relation to the SRS system will be developed.

1 Introduction
The 1st generation of the Danish SRS system (Noise reducing wearing course) based on CPX measurements according to [ISO/CD 11819-2] was developed in 2006. The SRS system has since 2006 been applied by both national and municipal administrations who are purchasing noise reducing wearing courses to a wide extent.

The 2nd generation of the SRS system is currently being developed due to old reference tires and wearing courses.

In the 1st generation of the SRS system, the noise was measured when the wearing course was newly laid, but not at any given time after paving. Together with the CPX measurements, the surface texture has been measured and analysed. This measurement is made to see how texture changes as well within the first months.

This project purpose is to obtain new knowledge of how the noise from the wearing courses changes within the first months of their lifetime, and to give a recommendation for when to perform noise testing of wearing course products to the SRS working group.

2 Wearing courses
The wearing courses were chosen from the SRS system and then constructed at the test site at road 145 located nearby Igelsø close to Holbæk. The daily traffic of the road is about 10.000 vehicles. The old wearing course of the road was before paving the new grinded up and repaved as a remix under the new wearing course.

Each of the wearing courses was paved over at least 500 m to have a sufficient length to measure with the CPX trailer.

The paving of all wearing courses was done from August 26th to September 1st where the first measurement was performed September 2nd.

A total of 5 different SRS wearing courses and a reference were paved at the test site, see [Table 1.]. The wearing courses are the most often used SRS types in Denmark.
Table 1. The chosen wearing course types on the test site.

<table>
<thead>
<tr>
<th>Wearing course type</th>
<th>Pictures of paved surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asphalt Concrete 8mm open graded</td>
<td></td>
</tr>
<tr>
<td>(AC 8o)</td>
<td></td>
</tr>
<tr>
<td>Stone Mastic Asphalt 8mm</td>
<td></td>
</tr>
<tr>
<td>(SMA 8)</td>
<td></td>
</tr>
<tr>
<td>Stone Mastic Asphalt 6mm + 8mm</td>
<td></td>
</tr>
<tr>
<td>(SMA 6+8)</td>
<td></td>
</tr>
<tr>
<td>Stone Mastic Asphalt 6mm + 11mm</td>
<td></td>
</tr>
<tr>
<td>(SMA 6+11)</td>
<td></td>
</tr>
<tr>
<td>Asphalt Concrete 6mm open graded</td>
<td></td>
</tr>
<tr>
<td>(AC 6o)</td>
<td></td>
</tr>
<tr>
<td>Asphalt Concrete 11mm dense</td>
<td></td>
</tr>
<tr>
<td>(AC 11d) (Ref.)</td>
<td></td>
</tr>
</tbody>
</table>

The measurements were made regularly after paving when the weather was good due to the standards. The noise measurements were made at 50 km/h and simultaneously with the texture measurement.

![Figure 1. The CPX/TEX noise trailer deciBellA from the Danish Road Institute.](image)

4 Results

In the following chapter, the absolute noise level results of the CPX measurement and some of the texture measurements are presented.

4.1 CPX results

The results of the CPX measurements are presented as the absolute noise level change in timelines for the wearing courses. As located on [Figure 2.], the noise emission of the wearing course types can be divided into 3 emission groups: The SMA 8, AC 8o and the AC 11d can be grouped, the SMA 6+ series can be grouped and the last group is the AC 6o.

![Figure 2. Timeline of the six wearing courses at 50km/h.](image)

The objective of this project is otherwise to locate when the noise emission from the wearing course...
beverages stable. An important calculation is then to determine at what the noise emission of the wearing courses are at day zero. For this, both logarithmic and linear approximation has been calculated as regression lines.

It was decided to present a linear regression of all noise emissions, where day zero is 0 dB, and then plot the change in noise emission for each measurement from the new reference zero. The equations for the regression lines are listed in [Table 2].

<table>
<thead>
<tr>
<th>Wearing course</th>
<th>Linear regression equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC 8o</td>
<td>Y=0.0214X-0.0294</td>
</tr>
<tr>
<td>SMA 8</td>
<td>Y=0.0234X+0.0425</td>
</tr>
<tr>
<td>SMA 6+8</td>
<td>Y=0.047X+0.3619</td>
</tr>
<tr>
<td>SMA 6+11</td>
<td>Y=0.0524X+0.3274</td>
</tr>
<tr>
<td>AC 6o</td>
<td>Y=0.035X+0.2895</td>
</tr>
<tr>
<td>AC 11d</td>
<td>Y=0.0365X+0.2895</td>
</tr>
</tbody>
</table>

The new time lines are shown in [Figure 3]. There are now 3 new groups of wearing courses that follow each other.

The SMA 6+ series is still the same, but the AC 11d and the AC 6o now groups and the SMA 8 and AC 8o groups.

As seen in [Figure 3] the change in noise emission from day zero to the last measurement day is from 1 to 2 dB and the largest change is from day zero until week 5.

This concludes the noise emission increases up to 2 dB within the first three to four weeks, and here after the noise emission flattens.

4.2. Texture results

In this sub chapter the texture data from the wearing courses which have the least and most change in noise is presented and discussed. The texture, from the SMA 6+11 and the AC 8o is presented. In each of the two figures, the results of the first (2-Sep) and the last (12-Oct) measurement is presented. The SMA 6+11 is shown in [Figure 4] where the spectrum decreases approximately 2 dB in all bands.

The AC 8o is shown in [Figure 5], where the spectrum increases approximately 1 dB in all bands above the 8 mm band.

In [Sandberg et al, p:414] there is a model for predicting noise, where the change in the 80 mm and the 5 mm octave bands are of interest. This model is also called the Sandberg-Descornet model.

The model says that an increase in the 80 mm band and a decrease in the 5 mm band should lead to an increasing noise emission and vice versa. 1 dB increase in the 80 mm band leads to an increase on 0.5 dB in the noise emission, where an
increase in the 5 mm band leads to a decrease of 0.25 dB.

In the AC 80 texture results, there is a small increase in the 5 mm but also an increase in the 80 mm band, which as measured gives an increase in the overall noise emission, in this texture measurement, the model correlates with the noise measurement.

In the SMA 6+11 texture measurement, the overall texture spectrum decreases 2 dB over time, which according to the model, should give a decrease in noise as well, this is however not the case since the overall noise emission increases 2 dB.

5 Conclusion

The CPX measurements of the wearing courses confirm that the measurement should not be performed within the first month after paving. The first measurement of 2011 of the pavement will show if more investigation of the early lifetime is needed. Further studies in the noise spectrum from the CPX measurement have to be launched to get a better understanding of the texture/noise correlation.

References

A Better Approach to Predict Subsidence due to Groundwater Leakage into Tunnels

Dr. Vikas Thakur
Senior Geotechnical Engineer, Geotechnical Division,
The Norwegian Public Roads Administration,
E-mail: Vikas.thakur@vegvesen.no
ABSTRACT

When groundwater (GW) leaks into a rock tunnel overlain by marine clay deposits or compressible sediments, it can cause a significant reduction in the pore pressure and induced more time dependent subsidence in the sediments. The effect of such subsidence is a serious issue with regard to the houses built around the tunnel areas. Subsidence analysis as adopted, today, for various road projects is based on classical consolidation theory. The increase in effective stress, which comes from groundwater-lowering (GWL), is assumed to induce an equivalent excess pore pressure that results in consolidation settlement. The settlements are then computed based on the classical theory of one-dimensional consolidation. However, in the particular type of problem, the total stresses in soil remains practically unchanged throughout the subsidence process. Therefore, excess pore pressure is not expected to be generated. In other words, subsidence calculation based on the consolidation theory is in fact not applicable for such type of GWL related problem. A real case example, the Hommelvik tunnel, from Norway is considered in this study to elaborate this. The paper presents subsidence data gathered during the past 20 years from representative locations at the tunnel site. The results obtained from the finite element (FE) analyses are discussed in light of these measured data and background assumptions. The FE analysis is seen to give a very promising framework to analyse such problems. This particular study calls for a better subsidence prediction procedure using a frame that also considers effect of creep.

Keywords: Ground water leakage, Subsidence, Creep, Finite element modelling

Research domain: Transport civil and road engineering
1 INTRODUCTION

When rock tunnels are constructed under populated areas, additional geotechnical challenges that could often be met are ground subsidence due to GW leakage into the tunnels. When GW leaks into a rock tunnel overlain by marine clay deposits or compressible sediments, it can cause a significant reduction in the pore pressure and induced settlements in the sediments. The subsidence associated with such pore pressure reduction could be significant especially in areas where there are reasonably thick layers of soft compressible clays overlaying the tunnels. The immediate effect of such subsidence is a serious issue with regard to houses situated around the tunnel areas. These subsidences are mainly results of an increase in effective stress, due to reduction of excess pore pressure, over a certain period of time. In addition, due to change in effective stress, the soil exhibits higher tendency to undergo time-dependent subsidence. Hence, stress and time-dependent settlements are expected to be induced due to the associate changes (Thakur et al., 2011).

The Norwegian Public Roads Administration (NPRA) report 103 published in 2003 suggests two alternative mechanisms associated with GW leakage into tunnels. Accordingly, a change in the pore pressure profile over a certain period may occur in two different ways as shown in Figure 1(A and B). Figure 1(A) shows a hypothetical situation where the GW level has reduced hydrostatically. This also implies that changes in the effective stresses are uniform along with depth. Another way is a gradual reduction of pore pressure at the bottom of a soil layer due to drainage towards the tunnel. This drainage is possible through open fractures created by the tunnel construction. This process is illustrated in Figure 1(B). In this particular type of pore pressure reduction, the GW level at the top remains unchanged whereas the pore pressure lowers at the bottom of the layer. Experiences from various tunnels in Norway suggest a combination of alternative 1 and 2 as the more likely scenario (NPRA, 2003).

The NPRA report also presents some data, from different tunnels constructed in Norway, regarding pore pressure reduction over and around the tunnels, see Figure 1(C). Measurements shown in the figure suggest that there is a correlation between the GWL and the distance from the tunnel. Hence, large GWL results in large influence area. For instance, NSB Abelhaugen had a maximum GWL of 13 m, and this resulted in 4 m GWL in about 450 m away from the tunnel. Additional factors which influence GWL are the natural GW flow...
patterns, local topography, soil permeability in horizontal and vertical directions, the rock fractures and thickness of overlying clay layers.

In this work, GW leakage induced subsidence is studied using a real case example, the Hommelvik tunnel, from Norway. The paper is divided into two parts; the first part of the study, presents detailed engineering characterization of the areas surrounding the Hommelvik tunnel. It also presents subsidence data gathered during the past 20 years from several locations at the tunnel site. This second part of the paper focuses on numerical analyses of measured subsidence and future predictions according to a guideline by NPRA and using a finite element tool. Primarily, this work presents a review of the papers by Thakur et al. (2011) and Degago & Thakur (2011).

Figure 1: (A): GW lowering (GWL) and pore pressure reduction due to leakage (Alternative 1). (B): Pore pressure reduction due to leakage without GW lowering (Alternative 2). (C): GWL with respect to the distance from a tunnel.
PART I: CASE DESCRIPTION OF THE HOMMELVIK TUNNEL

2.1 Case Description

The Hommelvik tunnel is located about 30 km east of Trondheim, Norway. The tunnel was constructed in 1989. The Hommelvik tunnel is a rock tunnel which was constructed under 10 – 12 m thick marine clay deposits. Today this tunnel constitutes one of the most important elements of the surface transportation between Trondheim city and the Værnes international airport. The Hommelvik tunnel is located under an urban area. Around 48 houses exist in the vicinity of the tunnel, of which 15 are constructed on clay deposits, 21 on rocks and the remaining 12 houses on combined ground conditions. Figure 2 shows an overview of the location of the tunnel and its surroundings.

Prior to the construction, settlement gauges were installed under several houses to monitor possible subsidence effects. To assess differential settlements several settlement gauges were installed on each house. Figure 2 shows some of the houses where subsidence have been monitored. The locations for pore pressure measurements, Hull 13 and Hull 16, are also

Figure 2. Overview of the Hommelvik tunnel.
shown in the same figure. Subsidence measurements gathered since 1989 shows that
subsidence are in between 35 to 150 mm. Some parts of the area are still under continuous
subsidence. The measurements show that the houses founded on clay deposits experienced the
largest subsidence; whereas, the houses partly founded on the clay deposit and partly on the
rock surface have suffered severe differential settlements. The Hommelvik tunnel resulted in
about 5 m of GWL. According to Figure 1(C), this may induce subsidence between 200 and
450 m away from the tunnel.

2.2 Predicted subsidence
Subsidence induced due to the tunnel construction was predicted a priori. Figure 3 shows that
the predicted subsidence is directly dependent on the thickness of compressible
layers (H) as well as the GWL. The worst case scenario occurs if the GW table is lowered to
the bottom of the draining clay layer. In such a scenario the subsidence that may occur due to
GWL, i.e. due to increase in effective stress, are presented in Figure 3. This was calculated,
using the equation proposed by Janbu (1970).

2.3 Measured subsidence and discussion
Immediately after construction of the Hommelvik tunnel, significant pore pressure reductions
as well as subsidence were recorded. The maximum pore pressure reduction, see Δu in
Figure1, was up to 60 kPa at Hull 16, which is on average equivalent to a 3 m GWL with
hydrostatic conditions. The immediate subsidence recorded on the area was around 50 mm.
At present, subsidence measurements are being monitored at 28 locations over the
Hommelvik tunnel area. In this paper, some representative results are presented. Some of the
locations were instrumented with more than one settlement gauge to observe the differential settlement. Subsidence measurements for the location B, is shown in Figure 4. Subsidence measurements at the other locations shown in Figure 2 are presented by Thakur et al. (2011).

![Figure 4. Measured subsidence at location B (ref. Figure 2).](image)

3 PART II: NUMERICAL MODELLING

3.1 The FE modelling

A finite element tool called Plaxis 2D v 9.0 is used in this study. The study is made with respect to measurement discussed in Section 2.3. A one dimensional oedometer condition is considered. The thickness of the clay layer considered for the analysis is 15.30 m. An axisymmetric model with very fine mesh, consisting of 15 node triangular elements, is adopted. Measurements showed that subsidence continue even in the locations where the GWL is expected to be minimal. This is considered to be an effect of creep. Besides, the soil in the tunnel area is characterized by its strong tendency to undergo deformation due to change of effective stress as well as time. Hence, the soil model selected for analysis is the Soft Soil Creep (SSC) model (Stolle et al., 1999; Vermeer & Neher, 1999). The SSC model is implemented, as a standard model, in Plaxis. The SSC is formulated based on early works by
researchers working on creep (Šuklje, 1957; Bjerrum, 1967; Janbu, 1969; Šuklje, 1969). Accordingly, the SSC model assumes a creep rate to be given by the current effective stress and the current void ratio (strain). In other words, any combination of void ratio (strain), effective stress and rate of change of void ratio (strain rate) is considered to be unique throughout primary and secondary consolidation phases.

3.2 Input parameters

Based on the soil parameters shown by Degago and Thakur (2011) the soil can be characterized with two layers. The top 2 m is a dry crust. The GWT is located at 2.2 m and the soil layer which is subjected to effective stress change is the layer from 2.2 to 15.3 m. Two incremental oedometer tests have been conducted within this depth, i.e. 3.7 m and 7.45 m. The tests indicate that the soil to be characterized by $\mu^*/\lambda^* = 0.04$. This ratio of $\mu^*/\lambda^*$ lies within the range for soft clays which is 0.03 to 0.06 (Mesri, 1987). In this study, a uniform soil layer is considered since the dry crust is not subjected to a change in effective stress resulting from GWL. The input parameters adopted for SSC model are given in Table 1.

<table>
<thead>
<tr>
<th>$\gamma$ [kN/m$^3$]</th>
<th>$k_v$ [m/day]</th>
<th>$\kappa^*$ [-]</th>
<th>$\lambda^*$ [-]</th>
<th>$\mu^*$ [-]</th>
<th>OCR [-]</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>2.4 e-4</td>
<td>0.009</td>
<td>0.083</td>
<td>0.0033</td>
<td>1.70</td>
</tr>
</tbody>
</table>

The symbols in Table 1 represent the following SSC parameters. $k_v$ is the vertical permeability of the soil; OCR (Over Consolidation Ratio) is defined as the ratio of preconsolidation stress to initial effective stress; $\kappa^*$ (the modified swelling index), $\lambda^*$ (the modified compression index) and $\mu^*$ (the modified creep index) are defined in Figure 5. Please refer the Plaxis manual (2011) for a description of the SSC model and the model parameters.

3.3 Calculation procedures

Settlement gages were installed on foundation of several houses, in 25/08/88, before construction of the tunnel. To be able to compare simulation results with measurements, three calculation phases are considered. The first phase is to simulate the condition before the construction of the tunnel, i.e. from 25/08/88 to 01/11/89. In this calculation phase, the soil creeps at its initial creep rate. The second calculation phase is to simulate the GWL as per the
idealization of NPRA guidelines discussed in Section 4. The GWT was lowered by 3.0 m within 345 days, i.e. from 01/11/89 to 15/07/90. The third calculation phase is a pure creep phase where the soil creeps at a relatively higher creep rate due to the effective stress change. This phase runs from 15/07/90 when the lowered GWT was assumed to stabilize till the end of measurement, i.e. 30/06/06. Another set of analyses were also carried out. In this analysis, one calculation phase is defined that simulate only a creep phase from 25/08/88 to 30/06/06 without change in pore pressure profiles. This analysis is done, in order to separate the effect of subsidence that is solely due to GWL.

3.4 FE results
The analysis result is presented along with subsidence measurements which were considered to be critical. The zero time in Figure 5 corresponds to a date of 25/08/88. In accordance, to Figure 5, construction of house could have affected the initial creep rate of the soil beneath the houses. In the analysis, this effect is considered to be negligible. In addition, water injection effect result in decrease in effective stress leading to possible unloading and creep at reduced rate. The effect of water injection is not also considered. Due to such simplifications, analysis results should not be compared directly with measurements. Hence, the predictions are separately presented (Fig. 5 upper) along with subsidence measurements (Fig. 5 lower).

As shown in Figure 5, the analysis result is in fairly good agreement with the measured subsidence despite the simplifications adopted in the analysis. It is also seen that with only creep effect, the subsidence is insignificant. However, when GWL along with creep effect is taken into account, the resulting subsidence becomes significant and the creep deformation with a creep rate that gradually decreases with progress of time. Hence, the overall predicted subsidence is in accordance to the measurements.

3.5 Prediction of further subsidence
Assuming that the calculated subsidence is representative of the measured subsidence history and the current conditions to remain unchanged, the next interesting step would be to predict future subsidence. The analysis and measurements in Figure 5,5 are shown for first 18 years. This simulation is continued for another 32 years, giving the settlement history of 50 years, Figure 6. Accordingly, the expected subsidence, within the next 32 years, is 50 mm. The settlement rate at the end of simulation is given by approximately 1.5 mm/year. This is lower than the creep rate predicted in 30/06/06 which was approximately 5.4 mm/year.
Figure 5. Subsidence measurements versus simulation.

Figure 6. Analysis for 50 years of subsidence history.
4 CLOSING REMARKS

This paper presents a lesson about how severe subsidence may occur due to pore pressure lowering induced settlement. The paper emphasized the importance of prediction and instrumentation to understand resulting subsidence over a large area due to tunnel construction. Predictions made prior to the construction of the tunnel were seen to underestimate measured subsidence. Moreover, subsidence are seen to continue with time and this calls for a better prediction tool that considers time effects as well as modifications on the current code of practice in Norway. The subsidence calculation using the SSC model in Plaxis considers two subsidence components, i.e. due to increase in effective stress as well as due to creep at higher creep rate. Several simplifications, discussed in the paper, are adopted in the analyses that could influence the results. Still, the FE analysis is seen to give a very promising framework to analyse such problems. Better problem definition and considerations would further improve accuracy of subsidence analysis. This particular study calls for a better subsidence prediction procedure using a framework that also considers effect of creep.

REFERENCES

Dejan Cerovac
Highway Institute
dejan_cerovac@yahoo.com

BRIDGE REHABILITATION
1. **Introduction**

It is widely considered that bridges are the most expensive part of the transport system. Irresponsibility, carelessness and lack of funds for maintenance, such as bridge repair, cannot be tolerated, as it is the case with the pavement structures, since the consequences may be serious – complete traffic break, a long-term repair.

The total number of bridge structures within the road network of the Republic of Serbia amounts to approximately 3000. Bridges are of different age and load carrying capacity, they were built with different material and they are exposed to different traffic volumes, so the need for a modern approach to all competent authority activities in the field of bridge engineering emerged, particularly in the area of management for existing bridges. Implementation of a Bridge Management System should be regarded as a modern solution for bridge management.

2. **History**

Construction of modern roads in the Kingdom of Yugoslavia before the II World War was followed by construction of bridges, at the very beginning of the period mostly in accordance with previous Kingdom of Serbia and Austro-Hungarian experiences, together with use of foreign, particularly French regulations, until 1929, when local bridge types were adopted. Since 1932, national loading codes were in use until 1949, when the well-known PTP-5 code came into effect. After the II World War the intensity of new road and bridge construction was significantly increased, using new material, for example prestressed concrete, but together with frequent neglecting of maintenance need. The condition of these bridges (i.e. its parts, when they are adapted) according to the data collected during database development, and with regard to age, is still considered satisfying. This is probably due to:

- Rigorous control during construction;
- Reserve in load carrying capacity;
- Cautiousness with processing of details.

On few occasions, detected damages appeared due to poor designs for widening and maintenance, especially on new parts, constructed after the year 1965.

In the Kingdom of Yugoslavia, in addition to already mentioned typical projects and regulations for road bridges, since 1930 there was a Law on Roads, and since 1935, a Decree on Cadaster of Roads and Bridges. All information on roads and bridges with detailed draws, along with information on detected deficiencies, have been entered into the Cadaster.
Until 1965 as regards to bridge construction on roads, reinforced concrete was the most frequently used, during the first years after the war the prewar types of Ministry of Construction were used, being abandoned after 1950, and since 1965, there is a large usage of prestressed concrete and prefabricated constructions. The construction period for several large bridges is probably from 1972-1976, when five large bridges were built over the rivers Danube, Sava, Drina and Morava. After the II World War, some legislation changes repeatedly took place, as well as changes in roads and bridges management, causing a negative impact on bridge condition, including those that have been built recently. Likewise, some activities such as bridge cadaster maintenance i.e. a thorough legally regulated monitoring of changes, are completely neglected and abandoned. Some positive changes have happened in this region during the eighties, especially after the adoption of Regulation on maintenance of main and regional roads. Deciding to found the Information Centre Department, as a part of the Road Head Office, the forming development activities of the Road Information System of Serbia started. The initial step in the process of its forming was the adoption of the Information System Rule. Adopting this Rule in 1992, the Road Head Office gained a very good base to start the formation of the Integrated Information System. The Bridge Information System development, as a part of the Road Information System, immediately started, but the first activities at the Bridge Data Base formation, started earlier in 1990.

3. BRIDGE MANAGEMENT SYSTEM

Bridge management system is consist of organizational elements, regulations, decisions, activities, standards, procedures and methods that help the road management to achieve the optimum results in bridge management activities within their competence, regarding technical and economic aspect, and in accordance with pre-approved load and budget constraints. This system refers to all activities related to each bridge, from the moment of its design and construction until the life end of a bridge, due to uselessness, uneconomical repairs, traffic inadequacy or some other factor. Typical management systems are based prevalingly on visual inspection and subjective assessment of structure behavior, therefore they are not considered sufficient for the maintenance of current structure condition nor for its capacity improvement, which is due to increasing growth of traffic volume. For this reason, the IT systems have been elaborated for structure managing, and they are currently available, allowing management that is more rational. Bridge Management System (BMS) is providing the optimization of the entire system maintenance through organized monitoring of bridge and structure condition, of maintenance costs, traffic control, and other significant elements and through rational short-term,
mid-term and long-term planning. For the system to function, it is necessary to develop a database on bridges and structures with updates based on regular periodic inspections. A certain amount of data refers to each bridge individually, and thus constitutes a bridge dossier. The rest of the data have the function in management system. Provided that the database had been supplied with all necessary details, it should have been able to show 3 information, required for the analytical system managing process:
- Costs (for management and users)
- Condition of bridges
- Anticipating a deterioration of bridges in the future

A simplified analytical process of management system starts with the approval of an authoritative and binding priority list of works, executed with the aim of improvement of bridges, which was obtained from the database, with closely related categories and quantities of works. In the next step, the plan and program are being created, followed by comparison with available means. If the means are sufficient, the realizations shall be performed by priority, and if not, the reduction shall be made.

Figure 1. The standard configuration for bridge management system

Other activities refer to design, construction, and the research work, and they are not included into the management system for the existing bridges, unless if necessary and to an appropriate extent.
Analyses, plans and programs refer to the following activities:

- Determining dynamics of database preparation and updating
- Planning dynamics for bridge examination, based on established intervals
- Preparation of short-term, mid-term, and long-term proposed plan and program for maintenance, repairs, rehabilitation, reconstruction, adaptation and replacement of the existing bridges
- Identification and analysis of the results of any previous planning period
- Timely notifying to the responsible authorities, institutions and public of the condition of bridges
- Planning the training of personnel for inspection and maintenance of bridges

In the area of information systems, the focus is on a database of bridges. The following is established:

- Structure of database
- Content of certain segments with required guidelines on collecting data
- The method of data storage, distribution and presentation
- Procedure for determining load carrying capacity and the issuing and recording permits for emergency transport
- Relation of centralized database and its possible decentralized parts

Monitoring of structure condition includes the following tasks:

- System of condition assessment
- The procedure and frequency of examinations
- Distribution of collected data
- Condition data operation

In the area of structure maintenance, the following activities are carried out:

- Analysis of the existing maintenance organization and improvement proposal
- Human resource analysis and improvement proposal
- Training of personnel
- Analysis of the existing equipment and proposal for procurement of advanced technologies
- Preparation of maintenance standards
The activities encompassed by the bridge management system are as follows:

- bridge inventory,
- maintenance and repairs of bridges (routine maintenance, repairs, rehabilitation, strengthening, reconstruction, adaptation and replacement of bridges),
- inspection and monitoring of condition (observation and monitoring of alterations (1 to 12 months), general inspection (1 to 3 years) and detailed (3 to 6 years), exceptional (as required), special inspection (as required)),
- inspection organization,
- inspection data operation,
- traffic monitoring,
- reports on accidents,
- exceptional heavy truck loads,
- cost computation.

The inventory will be the base for establishing the quantities for routine maintenance, so the segment with quantities of routine maintenance activities are the excerpts from the inventory by particular computation procedure, based on regulations on the maintenance of public roads and associate structures. Data record sheets on structure condition should include estimated repair works. A methodology for inspection and estimation of bridge condition has been established in 1992, and all bridges have been examined so far, using this methodology. The examination of a record sheet is providing the data on condition of 26 bridge parameters, with condition assessment on a scale of at least five grades, and with assessment of repair works quantity and category. The genuine procedure of the Highway Institute allows for creating ranking lists according to the overall bridge condition or any required parameter or set of parameters. This is currently the main priority indicator for bridge maintenance, and partially an indicator of load carrying capacity problems. Every bridge has a code associated with a route and a section, and provides with primary information on the type of obstacles, construction material and structural system. Likewise, it has its own identification number. A section identification number links a database on bridges with other databases.

The inspector (the condition inspector must hold a civil engineer diploma, must be experienced and trained for these types of works) is required to assess more precisely the types and quantities of repair works, if repairs are needed.

The basis for determining which bridges must be repaired is data on bridge condition contained in the database. Based on condition ratings for certain elements of bridge, and overall bridge
rating, and backed-up by an appropriate software application, the bridges to be repaired shall be
determined on the following grounds:

1. Ranking of bridge condition, which includes all elements of a bridge, rated both
structurally and functionally.
2. On the basis of extreme damage of individual structural bridge elements, whose condition
has been assessed as hazardous, (in Serbia, the scours and landslides in the area of
foundation are most frequent)
3. On the basis of insufficient load carrying capacity of the bridge, for everyday (usual)
traffic volume.

Bridges are the most specific parts of roads, built on locations where other solutions were
impossible or not cost-effective. Therefore, in terms of guaranteeing constant flow of traffic,
strict regulations must be enforced to them.

The main causes of damage to a bridge, by order of occurrence are classified as follows:

- Above ground: wind, atmospheric precipitations, temperature variations, smog, radiation
  and biological effects
- In the field: traffic loads, rain, snow and ice, flooding, flows of water and sediment, sea
  water and micro-organisms
- Within the ground: earthen materials, ground water with its constituents, landslides, tree
  roots, earthquakes.

In addition to security and usability, the concrete bridges should also possess the required
durability, which can be achieved by applying special measures in the design, construction,
operation and maintenance. Measures refer to preventing, reducing or slowing down the
degradation of concrete bridges, caused by the presence of water and moisture, aggressive
environment (chemical-salts, physical-temperature, and biological-vegetation), porosity and
insufficient thickness of protecting coating for concrete and appearance of rather large cracks in
concrete.

Durability of concrete bridges may vary and it will depend on design solutions, material and
structure quality, protection, natural conditions in the surrounding area, and maintenance quality.
The volume of works, whereby one provides for the conditions of safe and comfortable traffic
operation on the existing bridge during the long period is denominated as routine maintenance.
The purpose of routine maintenance is to prevent the appearance of damages, and to remove all
minor damages in the course of operation, as soon as possible.
From the diagram 1, it has been concluded that the omission of maintenance shortens the period of bridge operation, while occasional major repairs and strengthening ensure that the longest possible life cycle of bridge is achieved.

The Highway Institute’s Department for building materials is performing some tests in accredited laboratories, according to SRPS and EN standards approved to date, as well as controlled tests of relatively new materials applied during bridge repairs.

### 4. FIBER-REINFORCEMENT OF CONCRETE

Fiber-reinforced mortars and concrete are composite materials, which are obtained by strengthening of cement matrix, using evenly distributed fibers. The most commonly used types are steel, polypropylene, carbon, synthetics, glass, nylon, polyester, polyethylene fiber, etc. The reinforcement of cement matrix with steel fibers has contributed to the increase in toughness and ductility, as well as increase of flexural strength of underlying composites. Wide ranges of manufactured materials that are used for this purpose are synthetic materials similar to polymer.

When compared to steel fibers, the advantages of these fibers are significantly reduced shrinkage strain, reduction of water separation on the surface of concrete or mortar, easier plasticity and workability of fresh mixtures and lower cost. On the other hand, steel fibers have considerably better mechanical properties, particularly elasticity module, when compared to most types of polymer fibers. The advantages of concrete elements reinforcement with steel fibers are short period of works execution, compared to works performed with standard reinforcement, no need for conventional reinforcement, increased load bearing capacity due to load distribution,
increased resistance to dynamic loading, improved control of cracking, thus allowing for construction of plates with small thickness.

Testing of concrete reinforced with various quantities of steel fibers has been performed in the laboratory for concrete and binders, in accordance with the ASTM standard. Concrete prisms, 15 x 15 x 60 mm, were loaded with two concentric forces in thirds of span. As opposed to the situation in usual procedure of flexural strength testing, the growth of controlled deflection has determined the force growth.

Ordinary concrete structure breaks into two parts when it reaches the maximum tensile load, after which it cannot withstand any further loading or deformation. In fiber-based concrete, as opposed to ordinary concrete, the structure does not separate and is able to bear loadings until the appearance of quite severe deformations. Continuous loading i.e. energy intake, can contribute to the development of further deformations.

Due to their characteristics, concretes based on propylene fibers have quickly become applicable in the area of rehabilitation, in the situations when higher values of early compressive strength are needed and especially during the process of concrete placing in winter conditions.

Fiber-reinforced concretes are formed by application of fibers, and this provides for precise and easy installation; the risk of corrosion and reinforcement penetration on the surface has been avoided, and resistance to abrasion and impact has been increased. Tests have shown that polypropylene fibers maintain their original properties, even after six months of exposure to various aggressive impacts (high and low temperatures, salt waters effects, etc.)
Conducted experiments prove that mechanical properties are highly dependent on quantity of applied fibers (the usual dosage is 600-900 g/m³) and their length. The addition of propylene fibers contributes to the increase of mortar compression strength to a certain extent, and this effect is more prominent during the examination of younger samples (after 3 or 7 days). Following the application of these fibers, the increase of strength can be up to 20% higher compared to ordinary concrete within the first three days, while the final values may be higher by about 8%. Mortars with higher amounts of fiber have faster strength growth at early ages, as well as the highest values of compressive strength. As regards to fiber length, shorter fibers (up to 5 mm) enhance compressive strengths, whereas the application of longer fibers (over 20 mm) improves the deformation characteristics. Speaking of deformations, we must note that concretes based on fiber suffer 10-20% more strokes prior to the appearance of first cracks, and up to 2 times more strokes after the appearance of first cracks until the final fracture. The same properties are found in concretes supplemented by steel fibers in which the measured strain (deflection) is equal to strain in panels with dual reinforced mash. Early flexural strengths have been pronounced in a lesser degree, but full (final) strengths reach up to 30% higher values compared to ordinary concretes. Shearing strengths show up to 20% higher values compared to the standard. This phenomenon can be explained by the fact that in the initial stage of strengthening, cement matrix has insufficient strength, so that even very small amounts of propylene fibers (0.1 % of total volume) greatly contribute to the enhancement of this mechanical property. At later ages (after 7 or 28 days), cement matrix has a significantly higher level of hydration and proportionally greater strength, and consequently, the fiber contribution to the growth of composite strength is greatly reduced. Taking into account all positive aspects of fiber-reinforced concretes, we may easily conclude that their application is completely suitable for concrete repair. Polypropylene fibers were used during construction of pavement slab on the bridge Razanć, on highway E-75, section Belgrade-Nis, where it was inevitable to reconstruct the bridge slab with small thickness and required high mechanical properties.

Figure 1. Preparation of slabs for rehabilitation

Figure 2. Surface treatment of concrete slabs
5. CARBON STRIPS AND FABRICS

In addition to fibres being utilized as an additive for concrete, there are fibres meant to strengthen the concrete surface itself, so called FRP (Fibre Reinforced Polymer). The latter contains a great number fine fibres, with pronounced mechanical characteristics, within the framework of epoxide resin matrix. Depending on the type of fibre, there are aramid (AFRP), carbon (CFRP), and glass (GFRP) fibres.

The strengthening with carbon strips is applied in the cases as follows:

- Load-bearing capacity expansion,
- Damages, vehicle impact, fire or explosion,
- Requirement for higher safety ratio in the operation,
- When the walls are put out, or holes inserted onto the deck whereby the structural system is changed,
- Overcoming of design and execution errors, etc.

Carbon strips are made of fibres with diameters ranging from 0.01 mm to 0.1 mm, and may be supplied in the form strips – plates wherein the fibres are bonded with adequate epoxide binder-matrix and in the form of strips-fabrics obtained by the weaving of carbon fibre yarn.

The strips-plates can only sustain the loads (forces) in longitudinal direction, i.e. in the direction of its own strike. Their thickness ranges up to 2 mm, while their width reaches as much as 20 cm. The strips-fabrics can sustain the loads in a series of directions depending on the weaving of fibres:

- Monoaxial strips –fabrics have the fibres in one single direction,
- Biaxial strips have the fibres in two mutually normal directions
- Strips with diagonal weaving, where the fibres are at the angle of 45° in relation to the strip axis.

Their thickness is very small, i.e. several fractions of a millimeter, while their width is within the range from 20 to 100 cm. In comparison to steel plates, the carbon strips have several relevant advantages such as: smaller weight, very high tensile strength (even 8 times higher than the steel), small thickness, no potential corrosion whatsoever, easy processing and fitting-in, flexible behaviour, lower cost of placing.

Bonding of carbon strips onto the structural elements made of concrete is carried out by the application of adequate epoxide bonding glues delivered by the manufacturer together with the strips. Bonding procedures imply preliminary preparation of surfaces thereof, i.e. they have to be even or slightly corrugated with textured finish obtained by sand blasting or grinding. Department for building materials within the framework of The Highway Institute, Belgrade tested two-pack epoxide bonding glue utilized for mentioned purposes as regards the compressive strength and bending strength in accordance with Standard SRPS EN 196-1 (2008). The values obtained after 28 days, reached the strengths higher than 50 MPa, whereas the compressive strengths amounted to approximately 70 MPa. By utilizing the carbon strips, it is possible to avoid the breaking up of the structure and perform the repair and strengthening. That provides for good economic benefits since the works are carried out within the short timing, under every weather condition, without the interruptions of traffic in urban areas and on motorways, while the execution of works is commissioned to a small group of well trained operators. If it is necessary to repair the structure built earlier, then it is required to establish the diagnosis of such a cross-section.

The samples for strained edges of concrete are taken for this purpose by „pull-off“ methodology. The resistance to rupture test is performed on these samples, and if one establishes the bond stress to be higher than 1.5 MPa, i.e. the concrete element hardness to be higher than 30 MPa, then one may resort to the strengthening with FRP strips. Manufacturers provide the certificates from trial testing and detailed instructions for fitting-in. Besides, the designers have at hand the software back-up for reinforced concrete structures being strengthened, which is dimensioning the cross-sections of girders by adding steel reinforcements and carbon strips of specific sections onto the existing reinforcement. The testing of carbon fibre strips adhesion with the base (gripping power) in accordance with „pull-off“ methodology was performed on the basis of Standard SRPS U.M8.002 (1997) and technical specification received from the device manufacturer. This was done by bonding the circular dolly with standard glue and rapid-hardening mass (without dapping around the circular dolly) on the bridge cross girder. In both cases approximately identical average values were obtained pertaining to the pull-off stress (≈
6.5N/mm²) along with the remark that rapid-hardening mass was tested on the same day, while the glue, usually applied, was tested 7 days after. Upon the completion of „pull-off“ testing, the check-up of sound alteration was carried out by percussion hammer within the testing zone and in view that the sound did not change on circular dolly spot and within its vicinity, a conclusion was drawn that there was no segregation of carbon strips from the concrete.

![Fig.4 Carbon strips and panel on beam](image1)
![Fig.5 „Pull-off“ test on carbon strips](image2)

Woven fabric made of carbon fibres, glued with SikaDur 330 onto the concrete surface, was also utilized for the repair of bridges on main road M5 in Eastern Serbia in addition to afore-mentioned carbon panel strips. The gripping power with the base was tested by bonding the circular dolly as it was the case with panel strips.

### 6. Repair Mortars

Single-pack repair mortar bonded with cement was utilized for the repair of bridge across the river Crni Timok, improved with plastic mass which is usually used for the repair of concrete for surface correction and reprofiling of damaged parts in building and civil engineering construction. It is featuring resilient behaviour and could be processed either manually or with machinery. It is also performing perfect stability in the course of rendering it vertically or over the head along with high resistance to salts due to the frost. The material thereof was tested according to Standard SRPS EN 196-1 (2008) in order to establish its compressive and bending strength and thus determine the quality of repair mortar. Considerably higher values of bending strength for repair mortar were obtained in comparison with the concrete (approximately 14 MPa after 28 days), whereas the values of compressive strength reached even 80 MPa. Further works on the facility were granted with permission on the basis of results obtained by previous testing of repair mortar, as well as from „pull-off“ test carried out within the testing field.
7. **FINAL OPERATIONS**

Upon the completion of operations pertaining to the repair of concrete surfaces the same were coated afterwards. Group of EN Standards from 1504-1 to 1504-10 (Products and systems for protection and repair of concrete structures, specifications, quality control and harmonization assessment) defines the technical conditions for coatings which will be adopted soon by the Commission for Concrete and thus become compulsory. Concrete exposed directly to the carbonification, damaging impact of chlorides, abrasion of surfaces, impact of frost and other chemical agents can be protected by compounds for hydrophobic waterproofing and with coatings most frequently based on methyl-methacrylate and poly-urethane.

The compounds for hydrophobic impregnation set over the concrete surface, react chemically with the cement within the framework of concrete pores, thus obtaining silicon as final product. In such a way one prevents the capillary absorption of water and damaging materials diluted in concrete, yet the passage of water vapour is made possible through the processed surface.

The advantages of concrete coating refer to simple and fast setting by cold procedure, simple maintenance, high elasticity and capability of over bridging the cracks, high adhesion and full-fledged bonding with concrete base, optimal chemical resistance, high mechanical strength, short hardening period, optimal resistance to abrasion and skid resistance, water imperviousness and long-term protection. The concrete base has to be prepared properly with adequate evenness and has to take into account the weather conditions at the time of placing and one must admit that the cost is higher than those referring to hydrophobic impregnation. The protection of concrete surfaces with coating based on methyl-methacrylate (MMA) is composed of preliminary coating and MMA layer which could be set in the thickness either of 4 mm or 1.5 mm. If MMA layer has
the thickness of 4 mm, then it represents the waterproofing and wearing layer, with adequate skid resistance, while the layer 1.5 mm thick represents the backing coat resistant to skidding.

Concrete surface protection with MMA layer 1.5 mm thick does not include the protection coating in contrast to the above mentioned one, and the quantity of aggregate rubbing is somewhat smaller. Concrete surface protection with poly-urethane coating is carried out with layer 4 mm thick and represents the waterproofing and protective layer, resistant to skidding. Concrete surface protection durability depends on the coating product properties, condition and quality of concrete surface dressing, quality of works carried out, weather conditions during the performance of works, as well as depending on the conditions of how the treated surface will be exposed to them.

**LITERATURE**

[5] Sanacije, rekonstrukcije i održavanje betonskih konstrukcija u visokogradnji, Predavanja na gradjevinskom fakultetu
THE RELATIONSHIP BETWEEN ROAD DESIGN
AND DRIVING BEHAVIOUR

a Simulator Study
ABSTRACT

Speed is a substantial factor contributing to road safety since higher speeds lead to increased accident probability and severity. Young drivers are over-represented in accidents involving speeding. Currently, speed reduction is mainly achieved through law enforcement and the implementation of traffic calming measures. An alternative speed reducing approach would be to encourage drivers to choose the appropriate driving speed unconsciously. This could be done by clarifying the road design characteristics for each road category and thus making the road ‘self-explanatory’. Road width, horizontal and vertical alignment, roadsides and road markings are all identified as the factors contributing to road safety, especially in rural areas. It is important that road environment and design encourage good driving behaviour.

This paper aims to increase the knowledge of the effects of rural road design characteristics and the roadside environment on young drivers’ behaviour. This knowledge should help identifying which changes in road design, such as different cross sections and roadside environments, may lead to decreased driving speed.

The fixed-based driving simulator of DTU Transport at the Technical University of Denmark was used to conduct the experiment. Twenty young drivers aged between 20 and 26 years participated in the experiment. An existing stretch of the two-lane rural road in Denmark was used in the experiment. Road with no shoulders (existing road cross section), asphalt and gravel shoulders were tested. In addition, environments with and without trees along the roadside were applied. Thus, six experimental scenarios were obtained. The driving speed and the lateral position were used as performance indicators.

The results suggested that the simulator can be a useful tool to examine the influence of the road design on the driving behaviour, although several limitations of the fixed-based simulator have to be considered. Further, in accordance to previous results, there was a significant relationship between the road shoulder type and the presence of roadside trees on the driving behaviour. An introduction of shoulders provides greater homogeneity of the driving speed on straight sections and curves. As the drivers were adjusting their lateral position closer to the road edge when hard shoulders were present, the possibility of head-on collisions might have been decreased on the examined road stretch.

Although shoulder presence did not lead to decreased speed on the examined road stretch but still have a beneficial influence, it might be reasonable to examine the shoulder influence on the driving behaviour in combination with other design elements, for example, altered road markings in curves or rumble strips along the road edge and the centre line.
1. INTRODUCTION

In our everyday life everyone is concerned about road safety. Although many efforts have been undertaken to make roads safer, according to statistical figures 1.3 million road accidents a year cause 43 000 deaths and 1.7 million injuries in the European Union (Health-EU, 2011). The main objective of the Danish road safety strategy for 2012 is to reduce the number of fatalities and serious injuries in traffic by at least 40% in 2012 compared to the year 1998.

In Denmark one third of all road fatalities involve young drivers, even though this age group represents only about 10% of the licensed driver population (CARE, 2008). Young drivers, especially men, have the greatest tendency to overestimate their driving skills. Risky driving combined with a lack of driving experience is an important factor contributing to the high accident risk among young drivers (Gulliver and Begg, 2007). The risk of becoming involved in an accident is approximately 2.5 times higher within this age group when compared with all other male drivers (Clarke et al., 2005).

Additionally, for young drivers rural roads pose the greatest threat to their safety. In EU more than 3/5 and in Denmark 59% of the road accident fatalities amongst young people occur in rural areas, not on motorways (CARE, 2008). In general, with regard to the number of fatalities, rural roads represent the most dangerous road class for all drivers. 3/4 of fatalities on roads in Denmark happen on rural roads even if the accident rate on urban roads is considerably higher (SafetyNet, 2008). The results published by the Danish Road Directorate show that the mean speed on Danish rural roads, with a speed limit of 80 km/h, is 84 km/h and 14% of cars and vans exceeded the limit by more than 20 km/h (DRSC, 2000). There is considerable evidence that driving too fast for the prevailing road conditions, rather than high speeds per se, is a significant contributing factor to accident causation, although accidents at high speeds are more severe.

One of the greatest potentials for improving road safety lies in reducing speeds. Even though there is insufficient knowledge of how many people are killed or injured on Danish roads due to speeding, it is estimated that speed is the main cause in between 25% and 50% of all road accidents (DRSC, 2000). It was calculated that if the average actual speeds matched legal speed limits, the result would be 100 fewer deaths and 1500 fewer injuries each year.

Currently, speed reduction is mainly achieved through law enforcement and the implementation of traffic calming measures. Law enforcement only works if drivers actually know that they are speeding and still enforcement does not have long lasting effects. Studies show (Martens et al., 1997) that stricter penalties do not change drivers’ attitude towards the speed limit since they feel they are being forced to drive at a lower speed and do not choose this speed voluntarily. An alternative speed reducing approach would be to encourage drivers to choose the appropriate driving speed unconsciously. This could be done by clarifying the road design characteristics for each road category and thus making the road ‘self-explanatory’. Road width, horizontal and vertical alignment, roadsides and road markings are all identified as the factors contributing to road safety, especially in rural areas (Zakowska, 1997). It is important that road environment and design encourage good driving behaviour. In order to create an optimal road design for each road
type, information should be gathered on how road cross sections and roadside elements influence driving behaviour. In this study 7 hypotheses were established:

1. Participants will drive more slowly on curves than on straight sections.
2. Participants will drive closer to the centre line on curves than on straight sections.
3. The presence of shoulders leads to higher driving speeds.
4. Participants will drive closer to the road edge when a hard shoulder is present.
5. The presence of trees on the roadside leads to lower driving speeds.
6. Participants will drive closer to the centre line when there are trees on the roadside.

Only recently transportation engineers have found driving simulators to be powerful tools to test and evaluate the effects of new road layout, work zones, signs and signals, pavement markings, new construction and vegetation designs etc. prior to costly deployment. Furthermore driving simulation studies show that the driving simulator can be a viable tool to identify those at high-risk for speeding and for research on speed reduction measures (Pyne et al. (1995), Godley et al. (1997)) because of its high ecological validity and safety. So far it has rarely been used for this purpose. The use of questionnaires, a more common method for the prediction of risky driving behaviour, has been criticised because of its susceptibility to response bias (Ouimet et al., 2002).

2. METHOD

In the driving simulator study young drivers’ behaviour was examined on a rural road stretch. In order to achieve the most relevant results possible from this driving simulator study, it was decided to use a real road stretch as an experimental case. In accordance with the technical and visual abilities of the simulator, the stretch of Esrumvej was chosen (see Figure 1). Driving speed and lateral position was used as driving performance indicators.

2.1 PARTICIPANTS

Twenty one young drivers, 7 females and 13 males, were involved in this study. Participants were recruited at Technical University of Denmark. The participants were young drivers aged between 20 and 26 years (mean age 23.8, sd 1.58 years). All participants had a valid driving license, and their average driving experience was 4.9 years (sd 1.99).

2.2 APPARATUS

The experiment was conducted in the STSoftware simulator of DTU Transport. Simulator is composed of a cockpit model STS Jentig. It is equipped with all the necessary control systems. The graphics system consists of three 42” plasma displays. The front screen has 1920X1080 dpi
and the two side screens have 1360X768 dpi resolution. Displays are located around the cockpit providing a 180° horizontal and 40° vertical perspective. Rear and side-view mirrors are visible on the screens. Images are presented at a rate of 60 frames per second, creating the illusion of smooth movement. The visual objects are buildings, other vehicles, trees etc. Steering is performed using a steering wheel with force feedback. Furthermore, the driving simulator is equipped with a 5.1-channel 3D sound system, which provides the driver with the sound of the engine, wind and tyres thus enhancing the realism of the driving environment. All the systems mentioned can be controlled in a way to give the participant the most realistic possible driving experience.

2.3 EXPERIMENTAL ROAD: LOCATION AND GENERAL DESCRIPTION

Road Esrumvej (connects Helsingør and Frederiskværk) is located in the north of Zealand (see Figure 1) in Denmark. It can be classified as a middle speed distribution road. It is a two lane rural road (see Figure 1) In general the speed limit is 80 km/h, dropping to 70 km/h or 60 km/h on several stretches. All along the route the road has a varying lane width from 3 to 3.5 meters and there are no bicycle and pedestrian facilities along the road and at intersections. According to the traffic count on Esrumvej its annual average daily traffic (AADT) was estimated to 2.800-3.700 vehicles in 2003.

Figure 1: Esrumvej; Road section Skindersøvej - Hornbækvej
2.4 **ROAD SAFETY ASPECTS**

Esrumvej is reported to be fraught with several so-called black spots. A survey, held on Esrumvej during December 2002, indicated that according to drivers the main safety problems along the route were speeding, narrow traffic lanes, trees which are too close to the road edge and sharp bends (Wrisberg et al., 2005). As the road was designed and constructed several decades ago, it does not fulfill the safety requirements of nowadays. In order to follow the national plan to reach the goal of reducing the number killed and seriously injured in traffic accidents by 40% by the year 2012 (DRSC, 2000), safety improvements were made on several stretches of Esrumvej from 2001 to 2003. Improvements were made firstly by implementing rumble strips along the centre line. Secondly, where the road passes through villages, it was visually narrowed by green lines along both sides of the lanes (shown in Figure 3). Also, electronic traffic signs were implemented guiding the road users through the curves, together with electronic traffic signs indicating the allowed speed. The electronic signs ensure adequate visibility of the speed limit in poor weather conditions. In 2005 an evaluation of implemented safety features was carried out. Several pros and cons were highlighted. In general it has emerged that the average speed has been reduced by 0,2 to 6,2 km/h all along the route and that the number of accidents has decreased. On the other hand, noise has been increased due to the rumble strips, green road markings have faded away too quickly and electronic traffic signs are not arranged at a suitable angle for best visibility (Wrisberg et al., 2005).

![Figure 2: Esrumvej, 9.22 km](image)

![Figure 3: Section with green road marking](image)

2.5 **CHOSSEN ROAD STRETCH**

The chosen road stretch (Figure 1) is five kilometers long and the AADT is 3253 vehicles. The road is leading through the village Ny Horserød and the territory of the state prison in Horserød. The speed limit in general is 80 km/h, dropping to 60 km/h where the road passes through the villages. The road is characterised by many bends and for the most part there are single trees or rows of bushes on both sides. Predictability on the route is limited and the clearance area is reduced. The edge strip is narrow. Existing edges of pavement show sharp drop-offs (Figure 4).
In the case of a vehicle leaving the travel lane and entering the unimproved shoulder area, recovery becomes difficult and might result in the driver losing control.

Speed measurements in 2003 showed that the average speed on this road stretch is between 72 and 77 km/h. However, in some curves with such a narrow lane the speed even being belong the speed limits, is still too high for existing road geometry. On several road stretches speed registration data shows that 15% of drivers exceeded the posted speed (Wrisberg et al., 2005).

Road stretch goes through the forest with two sharp curves where recommended speed is 60 km/h. In total 39 accidents happened on this road section from January 2000 until October 2009. Most of them were run-off-the-road accidents (73%, 28) and had occurred in curves (60%, 23). The curve with most run-off-the-road accidents is shown in Figure 5. Also four head-on collisions on straight stretches and three collisions with roadside trees have been registered. Young drivers (age of 20 to 26) are involved in 28% (11) of the accidents.

![Figure 4: Sharp drop-off](image1)

![Figure 5: Curve where the most run-off-the-road accidents happen](image2)

## 2.6 ROAD DESIGN IN SIMULATOR

The experimental stretch was reconstructed in the simulator using on-site photos from the Danish Road Directorate home page (Directorate, 2011), Google maps street view (Maps, 2011) and a topographical map of the road.

Each of the test conditions had surroundings as similar as possible to real life (see Figure 6). Introduction of other variables that might affect the participants’ lateral position and speed was avoided.

There was occasional traffic going both directions. Traffic was more intense in villages (where data were not recorded) to concentrate drivers’ attention on the driving task more.
2.7 EXPERIMENTAL PROCEDURE

Prior to the experiment, the participants underwent a training phase, during which they were given the opportunity to familiarise themselves with the simulator. Afterwards six experimental sessions were carried out. Experimental conditions are depicted in Table 1. The visualization of each experimental condition is shown in Appendix A.

<table>
<thead>
<tr>
<th>Table 1: Experimental conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road without (existing) shoulders</td>
</tr>
<tr>
<td>with trees</td>
</tr>
<tr>
<td>without trees</td>
</tr>
<tr>
<td>Road with asphalt (hard) shoulders</td>
</tr>
<tr>
<td>with trees</td>
</tr>
<tr>
<td>without trees</td>
</tr>
<tr>
<td>Road with gravel (soft) shoulders</td>
</tr>
<tr>
<td>with trees</td>
</tr>
<tr>
<td>without trees</td>
</tr>
</tbody>
</table>

Each session was conducted over a five kilometres long road stretch. The cross-section of each type of the experimental road is shown in Figure 7. In all conditions, the experimental road had the same lane width (3.5 m). Shoulders are located on both sides of the road and they are 1.25 m wide. The centre line is doubled continuous, and the edge line is continuous for the road with hard shoulders and dashed for existing and soft shoulder conditions.

All participants were assigned to all experimental conditions randomly to counterbalance the conditions and therefore avoid so called ‘order effect’.

Each experimental session lasted for approximately four minutes and there was a short break between the sessions.
During each experimental session, the speed in metres per second and the lateral position in metres were recorded. The lateral position is measured as the distance between the center of the front bumper of the simulation car and the centre-line of the driving lane. If the center of the front bumper is to the left of this line, the value is positive and vice versa. The mean value of the lateral position states where exactly on the driving lane participants prefer to drive. A negative value means that the participant was driving closer to the edge of the road and a positive value indicates that the participant chose to drive closer to the center line. Increased weaving of the car (higher standard deviation of lateral position (SDLP)) represents a reduced vehicle control and may result in out of lane excursions.

The sampling frequency used was set to 10 Hz (data was recorded 10 times per second). The data set of each condition consisted of approximately 2200 observations, corresponding to 3 minutes and 40 seconds recording time. The measurements were done for curves and straight sections separately.

It has to be noted that several studies are carried out in order to compare speed and lateral position in the real-road and simulator environment. Blana and Golias (2002) stated that the mean lateral position and the SDLP is lower in real-road driving that in simulator driving. These differences may be mainly assigned to the fact that cues adopted by real-road drivers for distance perception are misused by simulator drivers and to the fact that the latter seem to underestimate the risk associated with roadside environment. The higher speeds in the simulator might be due to a lower perception of risk in the simulated environment, than would exist in the real road.

Due to a possible lack of speed perception in the simulator, each driver was maintaining a different mean speed, even though all participants were provided with the same instructions for the experiment. Therefore, the driving speed cannot be adjusted to the real-road environment.

The present study deals with matters related to the effects of fixed factors (see Table 2) rather than observing absolute speed and lateral position.
A Mixed-model ANOVA was employed to conduct an inferential statistical analysis. A significance level of $\alpha = 5\%$ is used. In the end, to compare the significance in the variations between shoulder types a Tukey-Kramer test was used to compare the pairs of group means.

## 3. Results

Results of the driving speed and lateral position are presented separately.

### 3.1 Driving Speed

The mean speed on curves is considerably lower than on straight sections (See Figure 8).

![Figure 8: Mean speed on straight sections and curves for each shoulder type](#)
The presence of trees does not influence the mean driving speed significantly (See Figure 9).

![Figure 9: Mean speed on straight sections and curves for each roadside type](image)

The null hypothesis that mean speeds on roads with different types of shoulders are equal, was rejected ($F = 6.59, p = 0.0017$). The highest mean driving speed is observed on roads with soft shoulders (3.53 km/h higher than on existing road), and the lowest on existing road (see Table 3). The mean speeds on curves were significantly different compared with straight sections ($F = 84.41, p < 0.0001$). The parameter estimate shows that speed was 7.34 km/h higher on straight sections.

**Table 3: Details of Mixed-design ANOVA model for mean speed**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Parameter estimate</th>
<th>numDF*</th>
<th>denDF**</th>
<th>F-value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shoulders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing: ref</td>
<td></td>
<td>2</td>
<td>214</td>
<td>6.59</td>
<td>0.0017</td>
</tr>
<tr>
<td>Hard: 1.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soft: 3.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Road</td>
<td>Curves: ref</td>
<td>1</td>
<td>214</td>
<td>84.41</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Straight: 7.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The number of degrees of freedom in the model; **- The residual degrees of freedom.

As it is shown in Table 4, variances were also found to be unequal between different road sections ($F = 78.44, p < 0.0001$). Participants varied their speed more on straight sections compared with curves. Also, it can be noticed that the mean speed varied significantly more on curves when there were trees on the roadside ($F = 2.77, p = 0.09$).
Table 4: Details of Mixed-design ANOVA for standard deviation for mean speed

<table>
<thead>
<tr>
<th>Factor</th>
<th>Parameter estimate</th>
<th>numDF*</th>
<th>denDF**</th>
<th>F-value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road</td>
<td>Curves: ref</td>
<td>1</td>
<td>217</td>
<td>77.36</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>Straight: 1.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trees</td>
<td>Trees: ref</td>
<td>1</td>
<td>217</td>
<td>0.33</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>No trees: 0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Road:Trees</td>
<td>Curves/Trees: ref</td>
<td>1</td>
<td>217</td>
<td>2.77</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>Straight/No trees: 0.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*-The number of degrees of freedom in the model; **- The residual degrees of freedom.

3.2 LATERAL POSITION

In general, participants drove closer to the centre line on curves (see Figure 10 and Figure 11). The mean lateral position on curves and straight sections for each shoulder type is depicted in Figure 10. There it can be seen that when a hard shoulder is present, participants drove slightly closer to the road edge whilst on a straight section. In all other cases the car is positioned closer to the centre line; more in the existing condition and less when a hard shoulder is present.

Figure 10: Mean lateral position on straight sections and curves for each shoulder type

When there are no trees on the road side, participants drove slightly closer to the road edge on straight sections, as it is shown in Figure 11. In curves, participants were driving closer to the centre line when there are trees on the roadside.
Figure 11: Mean lateral position on straight sections and curves for each roadside type

The null hypothesis stating that mean lateral position for each shoulder type is equal, was rejected ($F = 10.36, p = 0.0001$) (see Table 5). It can be seen that on average, participants were driving closer to the road edge when a hard shoulder was present and closer to the centre line in existing road condition. Similarly, mean lateral position for the road with trees along side was found to be different from the corresponding values for the road with trees ($F = 19.74, p < 0.0001$). Participants were driving closer to the road edge when there were no trees along the road. Also there was a significant difference between lateral position in curves and in straight sections ($F = 32.09, p < 0.0001$). Participants drove closer to the centre line in curves compared with straight sections.

Table 5: Details of Mixed-design ANOVA for mean lateral position

<table>
<thead>
<tr>
<th>Factor</th>
<th>Parameter estimate</th>
<th>numDF*</th>
<th>denDF**</th>
<th>F-value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shoulder</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing: ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard: -0.14</td>
<td></td>
<td>2</td>
<td>216</td>
<td>10.36</td>
<td>0.0001</td>
</tr>
<tr>
<td>Soft: -0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Road</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curves: ref</td>
<td></td>
<td>1</td>
<td>216</td>
<td>32.09</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Straight: -0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trees</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trees: ref</td>
<td></td>
<td>1</td>
<td>216</td>
<td>19.74</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>No trees: -0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*-*The number of degrees of freedom in the model; **- The residual degrees of freedom.

4. DISCUSSION AND CONCLUSION

This study showed an increase in driving speed when shoulders were present. This result can be explained by the results from a study done by Waard et al. (1995). In that study it was demonstrated that on roads with narrow paved width the mental effort needed to steer the vehicle
and to keep it in the relevant lateral position increased. Consequently, the drivers tended to reduce driving speed to decrease the mental workload accordingly. Based on that result, the increased speed found in this study when shoulders were present, might reflect a low level of mental workload allowing the drivers to carry out the driving task at a higher speed. The fact that the presence of shoulders generally leads to a higher driving speed, might inspire the conclusion, that the presence of shoulders decrease traffic safety. However, several studies (e.g. Zegeer and Council (1995), Ogden (1997)) stated that the shoulder paving was associated with a statistically significant reduction, even up to 41% in run-off-the-road and head-on collision frequencies on two-lane rural roads. One possible explanation would be that the extra space which was provided by the shoulders decreased the possibility of severe accidents even though the driving speed had increased. Additional space to correct mistakes in steering and avoidance of collision with oncoming traffic counterbalanced the increase of the driving speed. A momentary edge line crossing could possibly result in an accident when no shoulders were present, but it could easily be corrected in the case of present shoulders.

The result that the presence of soft shoulders increased the driving speed the most was unexpected. A possible reason of this might be the weakness of visual cues provided by the simulator (Kemeny and Panerai, 2003). It is possible that the difference in shoulder pavement in the simulated driving has not been obvious enough. Another option would be that soft shoulders were perceived as hard shoulders, just in a different colour. Therefore, it might be useful to introduce the post-experimental interview about shoulder types applied for each experimental condition in the experiment.

The results revealed that the participants drove closer to the road edge when hard shoulders were introduced. This might lead to a decreased possibility of head-on collisions. Moreover, a tendency was observed for participants to drive closer to the road edge when there were no trees by the roadside in this condition, therefore also run-off-the-road accident severity risk might be mitigated by the presence of hard shoulders.

As shoulder presence did not lead to decreased speeds but still might have a beneficial influence, it might be reasonable to examine the shoulder influence on the driving behaviour in combination with other design elements. For example, altered road markings or rumble strips along the road edge and centre line as it has been suggested by Davidse et al. (2003).

According to the data the trees did not influence the driving speed significantly. The drivers did not adjust to lower speed when there were trees along the road. The risk of severe run-of-the-road accidents including tree-condition is higher, as trees are by far the most commonly struck object type (Mok et al., 2006). According to the result of the present study, it can be stated that the trees by the roadside were not perceived as a threat to safety by drivers. Previous simulator studies by Horst, van der and de Ridder (2007) and Bella and Tulini (2010) revealed similar results.

On the other hand, the presence of roadside trees affected the mean lateral position, meaning that the participants chose to drive closer to the centre line, especially on curves, when there were trees by the roadside. This is in accordance with the driving simulator study by Antonson et al.
(2009). Also, by changing the lateral position closer to the centre of the road, the sufficient time and space margins around the driver might be maintained more carefully for the tree-condition. Sufficient safety margins are needed by the driver to feel safe and comfortable with no excessive mental load. This safety margin is used as a ‘Comfort zone’ by Cacciabue (2007). Concerning the lateral position, the presence of trees might mitigate the probability of run-off-the-road accidents.

Even though arboreal surroundings enabled drivers to adapt their lateral position further away from the road edge, the probability of run-off-the-road accidents still could not be found lower, as trees by the roadside did not cause speed reductions. Generally, higher speeds lead to higher accident risk as well as more severe accidents. With regard to the impact of roadside trees on safety, a compromise should be sought between the positive and negative effects. The results of this study indicate that the presence of roadside trees may serve to reduce the risk of run-off-the-road accidents, since the trees tend to cause drivers to drive closer to the centre line. However, since no reduction in speed was observed, the chance of serious injury or death is increased since collision with a tree becomes a risk. Although this study offers no conclusions regarding this compromise, it does highlight the dilemma faced by road designers as they attempt to improve road safety.

The results of this experiment revealed that the participants varied their speed on curves less when there were trees by the roadside. It could be argued that the trees might result in an increased mental workload which is then compensated by more stable driving. The reduction of speeding and the increase of homogeneity are known to lower the risk and severity of traffic accidents (e.g. Aart and van Schagen (2006), Goodwin et al. (2006), Nes et al. (2010)).

It can be concluded that an introduction of shoulders would not induce decreased speeds, but might provide greater homogeneity of speed on straight sections. Drivers are adjusting their lateral position closer to the road edge when hard shoulders are present, therefore the possibility of head-on collisions might be decreased on Esrumvej which was the test stretch used in the experiment.

The presence of trees by the roadside did not reveal significant changes in driving speed. The drivers did not reduce their speed when there were trees along the road, whereas participants chose to drive closer to the centre line when roadside trees were present. This might decrease the possibility of run-off-the-road accidents on the test stretch. It is suggested to do further examination to find the balance between roadside trees and road safety.

It can be stated that the results of this study concerning driving behaviour on rural roads with different road shoulders and different roadside surroundings can provide road designers with useful information concerning road safety.

The driving simulator can be a useful tool to visualise the impact of different road designs on the driving behaviour. The use of a driving simulator allows a better understanding of the physical space that is being designed without danger for the drivers. Also, quantitative evaluation of the safety of alternative designs can be examined, which enables the selection of the most appropriate before initiating real designs.
For the further research, it would be relevant to perform field study in order to validate the DTU Transport driving simulator. Field study could be done by measuring the vehicle trajectories with onboard devices to be able to link the simulator study results with the real life driving.
REFERENCES


Directorate (Danish Road Directorate). Internet source- http://www.vejdirektoratet.dk/roaddirectorat.asp?page=dept&objno=1024 (last access: 03.02.2011)


Health-EU (2011) Internet source: http://ec.europa.eu/health-eu/index_en.htm (last access: 03.02.2011)


Maps (Google Maps), Internet source: [http://maps.google.dk/](http://maps.google.dk/) (last access: 03.02.2011)


APPENDIX A

Illustration of each experimental condition

Existing road, with trees (1)
Existing road, without trees (2)

Hard shoulders, with trees (3)
Hard shoulders, without trees (4)

Soft shoulders, with trees (5)
Soft shoulders, without trees (6)
EVOLUTION OF PAVEMENT FRICTION:

Discerning the effects of load, velocity and aggregate types on polishing of pavements

Dan ZHAO*, Malal KANE, Minh-Tan DO, François DE LARRARD
French Institute of Science and Technology for Transport, Development and Networks, IFSTTAR, Route de Bouaye, 44341 Bouguenais, France
*PhD, IFSTTAR, Tel. : +33 2 40 84 57 17 ; E-mail : dan.zhao@lpc.fr (D.ZHAO)

ABSTRACT

Skid resistance of pavement is essential for the road safety. However, due to polishing, skid resistance decreases continuously during the whole pavement life. In all previous studies dedicated to pavement polishing, only the effect of heavy vehicles was taken into account until now because considering that, due to their low load, the effect of light vehicles was negligible. However, it was found a significant decrease in skid resistance on certain roads forbidden to heavy vehicles; therefore light vehicles (equivalent of the effect of different loads, velocities and materials) may well have a significant effect in the evolution of pavement friction too.

In this study, we try to quantify the effect of load, velocity and aggregate types on the skid resistance of pavements due to polishing. We consider that the parameters differentiating the two categories of vehicles regarding polishing are their loads and velocities. Via laboratory polishing test, the effect of these parameters and aggregate types during polishing tests is evaluated. The Wehner/Schulze test is used to perform the polishing process and to measure the friction. The obtained results show a real necessity to take into account the effect of light vehicle. Therefore, an update of a former model of evolution of skid resistance that took into account only heavy traffic is proposed to account now the effect of light vehicles.

Keys words: skid resistance, polishing, trucks, light vehicles, traffic, Wehner/Schulze
1 INTRODUCTION

Skid resistance of pavement is essential for the road safety. It is determined by many factors (velocity, contact pressure, environment, surface contact...). However, skid resistance varies during service life.

A programme of research was undertaken in IFSTTAR (French Institute of Science and Technology for Transport, Development and Networks) in France to identify phenomena such as binder removal, aggregate polishing by traffic to be responsible of these variations. Comparing skid resistance evolution of asphalt mix specimen and specimen of nude aggregate, the most significant point is that the aggregate and asphalt curves coincide after the asphalt has reached the maximum friction (see Figure 1). It can be said that the skid resistance evolution is controlled by the aggregates after the binder removal [Do et al., 2007; Tang, 2007; Kane et al., 2010; Zhao, 2010]. So this research programme concentrates on the effects of wear of aggregates on the evolution of skid resistance of pavements.

![Figure 1: Friction coefficient evolutions simulated by WS machine of asphalt and aggregate](image)

One of the main results of previous works was a polishing model that takes into account the effect of heavy traffic on skid resistance evolution. In addition, effect of different traffics is also identified to play significant and various roles on evolution of skid resistance, because of the difference of the mechanism of contact between the surface and the tire (contact area, pressure, velocity) [Kane and Do, 2007; Kane and Do, 2010]. Based on this mechanism and operation of the polishing process, this present work is taking into account the effect of different loads, velocities and materials.

The polishing system Wehner/Schulze is used to simulate the traffic polishing and to measure the friction during the polishing process. Different levels of these parameters are applied in order to quantify the evolution of skid resistance due to the application of a polishing process related to various traffic loads. Statistical analysis is implemented to identify the set of
factors, and an updated mathematical function is then proposed to describe the evolution of skid resistance.

2 EXPERIMENTAL PROGRAM

2.1 Wehner-Schulze machine

In this study, all measurements of skid resistance and polishing of specimen are performed with the Wehner-Schulze machine (WS-machine) (see Figure 2). This machines contains two stations for respectively performing polishing and measuring skid resistance.

The polishing station contains three rubber cones mounted on a rotary disc and rolling on the specimen surface. The rotation standard frequency is 500 rounds per minute, giving a linear speed of 17 km/h. The contact load between the cones and the specimen surface is 0.4 N/mm². A mix of water with quartz powder is applied on the specimen surface during the rotations. The surface is polished on a ring of roughly 16 cm diameter and 6 cm width. The machine can be programmed to stop after a given number of rotations. At each stop, water is projected on the specimen surface and 500 rotations are performed with the cones to wash all debris.

After the washing period, the specimen is moved manually to the skid resistance-measuring head. This head is composed by three small rubber pads (4 cm² area for each pad) disposed at 120° on a rotary disc. The contact load between the rubber pads and the specimen surface is approximately 0.2 N/mm². For the friction measurement, the disc is launched until a speed of 100 km/h is reached at its circumference. When the speed reaches 90 km/h, water is projected on the specimen surface. At 100 km/h, the motor is stopped and the disc is dropped until the
rubber pads touch the specimen surface. The rotation is stopped by friction between the rubber pads and the specimen surface; the friction–time curve is recorded. The friction value at 60 km/h is taken into consideration for analyses.

Load characteristics of the polishing station.
In the polishing station, two types of rubber cones with different size (normal cone (NC) and modified smaller cone (MC)) are used to simulate the load change (see Figure 3).

![Figure 3: Two types of polishing system cones of the machine Wehner Schulze](image)

(a) normal cone (NC)  
(b) modified cone (MC)

For some details of the average load in the contact area of each cone see Table 1.

Table 1. Description of the two types of rubber cone

<table>
<thead>
<tr>
<th></th>
<th>Normal cone (NC)</th>
<th>Modified cone (MC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applied load (N)</td>
<td>400</td>
<td>400</td>
</tr>
<tr>
<td>Apparent contact area (cm²)</td>
<td>111</td>
<td>76</td>
</tr>
<tr>
<td>Average load in the contact area (bars)</td>
<td>3.63</td>
<td>5.27</td>
</tr>
</tbody>
</table>

Velocity characteristics of polishing station.
Three types of velocity (200, 500 and 800 r.p.m) are applied for the experiment according to the possible range of the rotations of polishing head (from 200 to 1000 r.p.m)

2.2 Aggregate sample

Eleven types of aggregates characterized by their Polished Stone Value (PSV – see Table 2) and Micro-Deval (MDE) were used in this study.
Circular specimens are prepared in laboratory with 7.2/10 aggregate size (see Figure 4). They are fabricated by placing manually the aggregates in a single layer as closely as possible, with their flattest faces lying on the bottom of a mould, then filling the mould with resin [Do et al., 2007].

Table 2. Aggregate characteristics

<table>
<thead>
<tr>
<th>Aggregate</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Greywacke</td>
<td>Granite</td>
<td>Rhyolite</td>
<td>Rhyolite</td>
<td>Spilite</td>
<td>Diorite</td>
</tr>
<tr>
<td>PSV</td>
<td>60</td>
<td>58</td>
<td>57</td>
<td>55</td>
<td>53</td>
<td>52</td>
</tr>
<tr>
<td>MDE</td>
<td>10</td>
<td>9</td>
<td>11</td>
<td>9</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Aggregate</td>
<td>G</td>
<td>H</td>
<td>I</td>
<td>J</td>
<td>K</td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>Gabbro</td>
<td>Leptynite</td>
<td>Calcaire</td>
<td>Calcaire</td>
<td>Limestone</td>
<td></td>
</tr>
<tr>
<td>PSV</td>
<td>51</td>
<td>49</td>
<td>42</td>
<td>41</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>MDE</td>
<td>10</td>
<td>4</td>
<td>24</td>
<td>13</td>
<td>17</td>
<td></td>
</tr>
</tbody>
</table>

3 EXPERIMENTAL PROCEDURE

For each aggregate type one circular specimen was fabricated to investigate the material influences. The specimens were polished with the standard rotation 500 r.m.p and moved to the friction-measuring head afterwards.

Combining the aggregate characteristics (PSV and MDE), aggregates “F”, “H” and “I” were also used for the research of simulating the influence of different loads and velocities. The
specimen “F”, which have the medium level of aggregate characteristics, was chosen for the influence of velocities. The test program is shown in Figure 5.

Figure 5 : Aggregate experiment program

4 EXPERIMENT RESULTS

4.1 Influence of velocities

Figure 6 shows the evolutions of skid resistance of aggregate “F” with three types of velocity.

It can be seen that the difference of the three velocities is reasonably small between the tests, and the three curves coincide finally. As stated above it should be supposed that the velocity influence can be neglected.

4.2 Influence of loads
The evolutions of skid resistance of specimen “F” at different loads are shown in the Figure 7. The most interesting is, when the friction coefficients of bare aggregates discs versus the logarithm scale of the number of polishing are plotted, the succession of the experimental points is generally a straight line.

![Figure 7](image1)

*Figure 7 : Friction coefficient evolutions simulated by WS machine at two loads for specimen “F”*

The aggregate friction curves present a significant decrease of 0.182 for the specimen which is polished by the modified cone (dotted curve with hollow triangle), compared with the decrease of 0.165 for the specimen polished by the normal cone (curve with full triangle). This difference between the two loads can be seen also in the other two types of specimens (see Figure 8).

![Figure 8](image2)

*Figure 8 : Friction coefficient evolutions simulated by WS machine at two loads for specimens “H” and “I”*

It can be concluded that the effect of loads on skid resistance is not negligible and predominant from the aggregate component in this studied case.
4.3 Influence of materials

It can be observed in Figure 9, the group 1 compares the specimen with an average PSV between 55 and 60; group 2 shows the four specimen with PSV between 49 and 53; group 3 composes two specimen with PSV 42 and 41, and a single specimen K with the lowest PSV (32). The decrease is more significant for specimen “K”. For each group, the initial values are almost the same. However, after some polishing passes, each of them shows a different evolution tendency. The difference between polishing susceptibility is attributed to differences in their content of wear-resistant minerals [Bloem, 1971; Kowalski, 2007]. Limestone, specimen “K” with the lowest PSV 32 in group 3, is the most susceptible aggregate type to polishing, produces the lowest skid resistance [Csathy et al., 1968]. Aggregates of higher PSV may offer less decrease of skid resistance in general.

Figure 9 : Evolution of friction of bare aggregates versus the number of polishing passes
Figure 10 shows a reasonable correlation of friction coefficient versus the PSV value of aggregates after 180000 polishing passes. This study confirms perfectly a linear relation of PSV value versus friction coefficient of previous studies [Zhao et al., 2010].

\[ y = 0.0111x - 0.2543 \]

\[ R^2 = 0.8453 \]

Figure 10: Relation between friction coefficient and PSV after 180 000 polishing passes

5 MODIFYING THE KANE-MODEL

5.1 General

The resistance of an aggregate type against polishing is the key factor in providing skid resistance. The use of polish-resistant coarse aggregates or other aggregates with good frictional performance has always been considered as a useful way to maintain friction above an acceptable level [Bloem, 1971; Kowalski, 2007]. As it has been concluded that the skid resistance evolution is controlled by the aggregates after the binder removal [Do et al., 2007; Kane et al., 2010; Tang, 2007; Zhao, 2010], the skid resistance of an asphalt mixture pavement after a number of polishing passes could be simulated with the method of pre-evaluating like:

\[ \mu = (1 - d) \mu_B + d \mu_G \]  

With,

\( \mu_B \) is related to the type of asphalt, and is constant \((\mu_B^0=0.2)\) for asphalt mixture of aggregate “F” without ageing effect;

\( \mu_G \) is related to used aggregates, defined by aggregate function in this research;

\( d \) represents the surface fraction occupied by non bitumen covered aggregates on the total
surface, and:

\[ d(N) = 1 - \exp\left( -\frac{N}{N_B} \right) \]  

\( N_B \) is related to the number of polishing passes at which all the asphalt over the aggregate is completely removed. This parameter can be taken equal to 5000 for asphalt mixture of aggregate “F”.

In this research, it has been observed that the aggregate function can be defined by the influence of loads and materials. According to a simple linear relationship between polishing, load and velocity defined by Kane [Kane and Do, 2007], we propose an aggregate function \( \mu_G \) to take into account the load and material effect as following:

\[ \mu_G(N) = -a \cdot \log(N) + b \]  

With \( N > 0 \), is the number of polishing passes.

Where, we suppose that “\( a \)” represents the rate at which aggregates lose friction during the polishing passes, and the value can be linked to aggregate PSV; “\( b \)” is a constant, and can be determined by the load.

5.2 Identification of “\( b \)”

In this research, we have simulated two types of loads (normal cone and modified cone) by changing the rubber cones size to provide adequate frictional resistance at various average traffic levels. Using the approach of wear law of Kane [Kane and Do, 2007], which considers that the material removal due to polishing is proportional to the applied load and to the relative velocity. We supposed that “\( b \)” in the aggregate function is proportional to the applied load between the polishing head and the specimen (Figure 11).

\[ b = -b_1 \cdot P + b_2 \]  

With \( P \) is the average load in the contact area; \( b_1=0.0167 \), and \( b_2=0.6445 \) in this research, defined by the fitting of experiment of the three aggregates (F, H, I) with these two types of loads (normal cone and modified cone).
Figure 11: Identification of “b”

Figure 12 shows the experimental results and model prediction of the evolution of skid resistance of bare aggregates versus the number of polishing passes, using the parameters in Table 3. However, the main issue in this equation is at N=0 where the logarithm becomes infinite. Instead of using the above aggregate function, we define that the initial value of $\mu_G(N=0)$ in the model is the initial value of experiment.
The value “b” defined by the above model can be used to plot the aggregate function of others aggregates with the standard load (normal cone). And it is been supposed that “b” is constant for all the aggregates which correspond the same load (see Table 3): $b=0.584$ for normal cone and $b=0.555$ for modified cone. In this paper, there are only three aggregates which have the fitting value of modified cone.

Table 3: Parameters used to plot the aggregate function

<table>
<thead>
<tr>
<th>Aggregate</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>G</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>0.029</td>
<td>0.044</td>
<td>0.031</td>
<td>0.020</td>
<td>0.057</td>
<td>0.048</td>
<td>0.049</td>
<td>0.045</td>
<td>0.077</td>
<td>0.079</td>
<td>0.096</td>
</tr>
<tr>
<td>$b$</td>
<td>NC</td>
<td>0.584</td>
<td>0.584</td>
<td>0.584</td>
<td>0.584</td>
<td>0.584</td>
<td>0.584</td>
<td>0.584</td>
<td>0.584</td>
<td>0.584</td>
<td>0.584</td>
</tr>
<tr>
<td></td>
<td>MC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.555</td>
<td>0.555</td>
</tr>
</tbody>
</table>

5.3 Identification of “a”

The fitting value of “a” is shown in Table 3. The scope of “a” is from 0.020 to 0.096 for the PSV of aggregates respectively from 32 to 60, with “b” constant of 0.584 for the polishing passes of normal cone and of 0.555 for the modified cone. The results shows a linear relation between “a” and PSV (see Figure 13). The continuous curves correspond to the fitting linear trends as a function of material like:

$$ a = -a_1 \cdot PSV + a_2 $$

With $a_1=0.0025$, and $a_2=0.1787$ in this study.

![Figure 13: Relation between “a” and PSV with “b” constant for one type of load](image)

There is reasonable correlation between the aggregate PSV and the parameter “a”. The result should be further investigated with additional experimental data.
5.4 Validation

Figure 14 represents the experiment (the points) and model data (the lines) of the evolution of friction of different aggregates, Figure 14d plots the evolution of friction versus the number of polishing with logarithm scale. Aggregates friction starts from a maximum then decreases. It can be observed that the skid resistance coefficients of function versus the logarithm of the number of polishing fit well for most of aggregates.

Figure 14: Comparaison between experiment and model, evolution of friction of bare aggregates versus the number of polishing passes

In Figure 14d, it can be seen that the line slope defined by “a” is various for different materials, but the intercept “b” of the friction coefficient is the same for all the aggregates because of the same used load.
5.5 Prediction of evolution of skid friction

If we apply 3 bars of the apparent load for light vehicles and 11 bars for trucks [Hamlat, 2006; Michelin, 2000], we could calculate and simulate the wear effect on the aggregates under these two different loads (see Figure 15).

![Figure 15](image)

*Figure 15: Comparaison between light vehicles and trucks effects: friction evolution of bare aggregates versus the number of polishing passes*

With the above aggregate function and their parameters, the skid resistance can be determined according to the asphalt function (equation 1).

![Figure 16](image)

*Figure 16: Comparaison between light vehicles and trucks effects: friction evolution of asphalt mixes versus the number of polishing passes*
Figure 16 show the prediction of evolution of skid friction versus number of polishing of two loads (light vehicle and truck). Two parts of asphalt curves can be confirmed from the simulated evolution tendencies of the asphalt mixture, the friction first increases to a maximum value, then decreases with further passes of polishing, and coincide with aggregate curves.

6 CONCLUSIONS

This study provides a set of experimental data concerning the evolution of skid friction of nude aggregates under the effect of load, velocity and aggregate types. The scope of the study then was extended to the effect of traffic, equivalent of different loads and velocities. A physical model for predicting the evolution of friction taking into account the material and load parameters is proposed. A proposed aggregate function is derived from the Kane hypothesis, which confirms that friction on the polished surface is proportional to the applied load and the material coefficient. The material coefficient is supposed to be proportional to the aggregate characteristic PSV. Additional experimental data are needed to further investigate the effect of load on evolution of skid friction.

7 REFERENCES

Kowalski, K.J. (2007). Influence of Mixture Composition on the Noise and Frictional
Characteristics of Flexible Pavements. Ph.D. Dissertation, Purdue University, West Lafayette, IN.
SUSTAINABLE MAINTENANCE OF RURAL ROADS IN SLOVAKIA

Authors:
1. Ing. Lubomír Pepucha, PhD., Assistant of professor, University of Zilina, Faculty of Civil Engineering, Department of Construction Management, Slovakia
   lubomir.pepucha@fstav.uniza.sk
2. Ing. Martin Pitoňák, PhD., researcher, University of Zilina, Faculty of Civil Engineering, Department of Construction Management, Slovakia
   martin.pitonak@fstav.uniza.sk

TOPIC:
Transport civil and road engineering

KEYWORDS: sustainable maintenance, Pavement management system, economic effectiveness, cost optimization,
1. **Introduction**

Road management in Slovakia is still one of decisive instruments of the Slovak economy. Road transport is the most important and the most frequently used component of transport infrastructure in Slovakia and at the same time is one of the key factors of growth in modern economies, towards which the Slovak economy is approaching with its growth. This fact even more enhances the need for continuous innovation of all support tools of transport development and transport infrastructure. Investments in infrastructure are considered to be an effective way to stimulate the economy and initiate the growth trends. Satisfactory condition of the road and highway network is therefore even today one of the all-society priorities, essential for the development of society and the state. In the European Union (EU) is the support of transport infrastructure considered a part of strengthening the society integration and creates a positive development throughout the region concerned. Currently, the EU seeks to intensify support for the development of transport networks. Therefore, organizations responsible for administration and operation of the road management of the Slovak Republic fulfil an important and at the same time an exclusive role. Following the restructuring in this sector these organizations are:

1. National Highway Company - highways and expressways
2. Slovak Road Administration - 1st class roads
3. HTU - Self-governing higher territorial units - 2nd and 3rd class roads.

![Road Network in the Slovak Republic](image)

*Figure 1: The Slovak Republic road network according to transportation importance [5]*
Roads are part of the Slovak Republic transport infrastructure. There are highways, expressways and roads divided into three classes. Some roads are also part of the Trans-European North-South Motorway Project (TEM) and the International E-road network, designed and created by the United Nations Economic Commission for Europe.

Various analysis and evaluations of the efficiency of public benefits from the road transport in recent years have highlighted the fact that the amount of funds expended on the development and maintenance of the road and highway network, as the top processes in the road management, does not reflect actual requirements. Organisations managing the road network therefore feel the need to improve procedures and methods for determining the efficiency of their operations and thus use the allocated resources more economically and efficiently.

The paper presents examples of alternative practices that have been developed in Slovakia to achieve sustainable maintenance of rural roads (roads void of highways and expressways). The maintenance of rural roads has been a relatively lower priority in funding in last two decades, despite the rural road network comprising a large proportion of the overall road network in Slovakia. This has resulted in significant deterioration in the technical condition of the rural road network. It was reason for Pavement Management System development. The goal is to reach increased level of technical condition of rural roads on the level of road network, level of services and accessibility to road users and of course to increase the safety on the level of road network. The paper presents comprehensive system developed in Slovakia for Sustainable Maintenance of Rural Roads. System consists from road network monitoring, road construction diagnostics, traffic volume prediction, design of the repair and maintenance technologies, using of software tools for economic effectiveness evaluation. The system is applied every year for selected road section chosen on the level of road network by the local road administrators according surface condition. The inputs, which are taking account are non-variable parameters (horizontal alignment, vertical alignment, road class …), variable parameters (bearing capacity, residual lifetime, roughness, skid resistance and surface condition), traffic volume and its prediction, designed repair and maintenance technology, investment costs. The system output is the sequence made according the economic evaluation of prepared investment projects to repairing or to maintenance technology application. Economic effectiveness evaluation is founded on comparison of investment costs with road user benefits lead from road section repair and maintenance works realization. The final step is that every year Slovak Road Administration allocates the budget indented for repair and maintenance works for road network according the sequence founded on objective economic criteria. This system ensures the optimal allocation of the limited budget. The allocation is founded on objective criteria and data result from using of up to date methods and tools for design, diagnostics, maintenance and economic effectiveness evaluation in the decision making processes.

The maintenance of rural roads has been a relatively lower priority in funding in last two decades, despite the rural road network comprising a large proportion of the overall road network in Slovakia. This has resulted in significant deterioration in the technical condition of
the rural road network. Unexpected and sudden increase of traffic volume on the roads, caused by general expansion of capital construction and development projects in recent years, incurred that road structures, not designed for such traffic load intensity, damage much faster than planned. Thus the need for financial resources to maintain and repair the roads of all classes is rising proportionally. Public resources allocated for pavement maintenance and rehabilitation Works are insufficient and budgets of road administrators are undersized. This means that in Slovakia, like in other countries, it is necessary to consider all options to obtain more effective and efficient management for road systems.

The next aim of this paper is therefore to focus on key activities of road administrators and to advance the principles of process management with emphasis on effectiveness of top processes in this field. This paper is focused on top process of using the road, representing the longest and the most important life-cycle phase of the road. The subject are key activities of road administrators (basic processes) aiming to increase the efficiency of their activities by measuring their performance, evaluation and proposing objectification of allocating the resources available for maintenance and repairs of the road network to individual administrators.

The task is to present functional and at the same time universal model of evaluation of top processes in the road management based on measurable performance indicators or quality processes which will be applied to one type of administrator (Slovak Road Administration), but will also be applicable to any of the entities operating in the area of the road management. The presented methodology will improve the efficiency of processes of road administrators by evaluation of effectiveness of processes based on quantitative assessment and interpretation of their measurable indicators, and also allow comparison with other organizations involved in the management and maintenance of roads aiming for their own improvement.

2. **Pavement Management System**

There was a significant need to develop a management system for road network in last two decades because of degradation of the road network pavement constructions. The goal is to reach increased level of technical condition of rural roads on the level of road network, level of services and accessibility to road users and of course to increase the safety on the level of road network. The paper presents comprehensive system developed in Slovakia for Sustainable Maintenance of Rural Roads. System consists from road network monitoring, pavement construction diagnostics, traffic volume evaluation and prediction, design of the repair and maintenance technologies, using of software tools for economic effectiveness evaluation. The system is applied every year for selected road section, chosen on the level of road network by the local road administrators according surface condition. The inputs, which are taking account are non-variable parameters (horizontal alignment, vertical alignment, road
class …), variable parameters (bearing capacity, residual lifetime, roughness, skid resistance and surface condition), traffic volume and its prediction, designed repair and maintenance technology, investment costs. The most important activity is selection of the road sections for the investment to repair and maintenance works.

2.1 SELECTION OF THE ROAD SECTIONS

The basic prerequisite for the success of the PMS application is the selection of road sections. At the renewal and modernization of the road network in the Slovak Republic is currently being used the Pavement Management System (PMS). PMS is the process pursuing the effective use of pavements of the road network in given sections, in certain operating conditions, involving systematically organized maintenance, repairs and rehabilitation of pavements, in terms of the most economic use of financial, material and energy resources.

On the basis of diagnostic measurements and technological possibilities of repairs are by means of optimizing and economic criteria selected sections of roads in accordance with the order of urgency. The process itself is shown in Fig. No.3. Final allocation of available financial resources for particular sections of the road network is performed on the basis of economic efficiency, expressed by the percentage of the internal rate of return from the costs of proposed rehabilitation technology. Currently is used so-called PMS - preferred, which is based on the current technical condition of pavement.

The Slovak Road Administration as owner and administrator of the road network request every years all local organizations of the road administration to select, on the basis of visual inspections and assessment of pavement surface conditions, sections which require rehabilitation investment. These sections were subsequently subjected to a thorough diagnosis of technical conditions of the road, implemented by the National Road Databank. Department of Pavement Diagnostics collected the data on pavement operational capability and performance parameters better said pavement efficiency parameters/ skid resistance, roughness, surface condition and pavement distress - bearing capacity/ of sections of roads proposed for rehabilitation. The processing of accumulated data and determination of the order of importance of repairs based on the internal rate of return (IRR) was also processed by the National Road Agency. Diagnostics/Pavement evaluation/ consisted of assessment of all relevant variable pavement parameters: [2]

- Bearing capacity - Kuab FWD 50: bearing capacity and residual life of pavement construction [9], [14].
- Roughness - Profilograph GE: - The average longitudinal and transverse unevenness/mean rut depth, IRI [9], [14] [10].
- Surface condition – Videocar: - surface condition expressed by Pavement Serviceability Index [9], [14].
- Skid resistance: Skiddometer BV 11: - skid resistance measurement [9], [14].
PMS on the basis of the described method determined the priority order of urgency for rehabilitation of roads. It turned out the selection of the road sections of all rural road network in Slovakia in the state of disrepair, which require immediate rehabilitation.

Figure 3: PMS as a transformation system represented by the block scheme[1].

where:

PPV - input variable parameters describing the state of the pavement in terms of PV and PS
PV - Operational Performance
PS - Operational Capability
ZZ - Residual Life
HZ - Thickness of Strengthening
ON - Cost of construction work
NSÚ - Cost of continuous maintenance
UN - User Costs
EN - External Costs
SP - Social Benefits
NPV - Net Present Value
CANUV - Data obtained by measurements on pavement by the CANUV system
KT - Library of Technologies
DP - Transport parameters
2.2 DESIGN OF THE REPAIR AND MAINTENANCE TECHNOLOGY

The term itself "the Pavement Rehabilitation" is in our case understood as a pavement strengthening - on average 40 mm new asphalt layer, in order to secure pavement operational capability as well as operational performance / pavement efficiency/ for a period of 10 to 15 years. [8]

The economic efficiency evaluation of costs of rehabilitation PMS of selected sections was carried out by ISEH (Integrated System for Economic Evaluation) developed by the University of Zilina in cooperation with the Slovak Road Administration. [7]

The program calculates the Road user costs (RUC) before investment and compares it with RUC after investment. (time delaying or saving, Vehicle operation costs – fuel and oil consumption, tires degradation, maintenance costs ) The outputs are Payback Period (PP), Net Present Value (NPV) and Internal Rate of Return (IRR). This software uses the SRA every year on the level of road network for decision making process. Based on demands of local road administrator, economic effectiveness of planned maintenance works on selected sections, SRA made the rank of road section for evaluated year. The main criterion is IRR and it does decide which of the planned investment to maintenance works will take place. [6]

2.3 ECONOMIC EFFECTIVENESS EVALUATION

The system output is the sequence made according the economic evaluation of prepared investment projects to repairing or to maintenance technology application. Economic effectiveness evaluation is founded on comparison of investment costs with road user benefits lead from road section repair and maintenance works realization. The final step is that every year Slovak Road Administration allocate the budget indented for repair and maintenance works for road network according the sequence founded on objective economic criteria. This system ensures the optimal allocation of the limited budget. The allocation is founded on objective criteria and data result from using of up to date methods and tools for design, diagnostics, maintenance and economic effectiveness evaluation in the decision making processes. The main advantage is that all processes are transparent and if the system is continuous and every year updated, the financial resources for the systems are at the accepted level and all processes of the system have software support.
3. **APPLICATION OF THE SYSTEM IN THE PROGRAM INTERFACE**

Presented part of Pavement management system (PMS) is applied into program interface called: Integrated system for economic evaluation of the pavement building repair technologies (ISEH). Program ISEH Software ISEH (Integrated System of Economic Evaluation) has three modules: Basic, Preferred and Optimizing. Preferred and optimizing modules are based on basic modules.

Preferred module works with data about present technical condition of the road construction and road pavement, present parameters of the operating ability and operating capacity of the pavement construction. This module enables to make sequence of the road section from the view of the repair works realization necessity for each valuated road section on the level of road network. This sequence is based on designed building repair technology economic effectiveness evaluation, its purchase costs and external-social benefits.

Optimizing module works with data about present pavement technical condition and future progress of the pavement construction technical condition. Next module works with present parameters of the operating ability and operating capacity of the pavement construction and its prediction based on predicted functions for all variable parameters of the pavement construction. This module enables to make the sequence present urgency of the repair works for all valuated road sections. This sequence is based on designed building repair technology economic effectiveness evaluation, its purchase costs and external benefits during
the technology life time. This sequence can be updated for various years according changes in operating ability and operating capacity of the pavement construction, changes in designed repair technology, its price and changes in external - social benefits.

Application of the PMS economic part for economic evaluation of the repair technology design has introductory window that enable motion between single parts of the program application and enable to vote between all modules of the program application.

3.1 Inputs

Ranges of inputs depend on type of the program module. You can work with basic, preferred or optimizing module. In principle it takes the data: identification data, data about technical condition of the pavement and pavement construction: traffic parameters (traffic volume, traffic flow structure ...), variable parameters of the pavement surface, residual life time – variable parameters, road geometry (unchangeable parameters) etc…

![Figure 5: Structure of the user interface](image)

3.1.1 Identification data

Identification data include common data about evaluated road section. (Name, type of road, road ID etc..), economic data (year of project realization, repair works investment costs, residual life time of the pavement construction before and after project realization, bank rate – discount of the Slovak National Bank etc..), data about road section localization from the view of the Slovak road network – data from Nodal Location System (LNS) (nodal points, length...
of the road sections – this application (work in LNS) is possible only if the user choose the work in LNS by the creating of the new action) and unit prices of the road user and external costs.

3.1.2 Interest of the road users (Traffic volume)

Traffic volume data are entered like table of 24 hour traffic volume for all traffic flow structure for evaluated road section. If the user choose by the creation of the new action possibility enable enter the data for each traffic flow direction independently, user have to enter the percentage of the traffic volume in the direction of the road stationing. User has to enter the percentage of the traffic growth – traffic growth coefficient.

3.1.3 Operating ability (variable parameters)

Operating ability present the immediate conditions of the pavement represent by the values of the variable (mutative) parameters of the pavement construction: longitudinal roughness, transverse roughness, surface condition, skid resistance etc.. All of the parameters are entered for the all length of the evaluated road section. According the new action setting selected by user, user can enter the data for both direction of the road section or for each traffic flow direction separately. In the second case is on the bottom of the window displayed button for changing the traffic flow direction and copying from one direction to the reverse.

3.1.4 Geometry and pavement construction data (Unchangeable data)

These data include: Road category, road type, road geometry – road alignment, traffic interruptions (crossroads), limited traffic speed, urban areas etc…

In the case of the Optimizing module user enter the residual pavement construction life time too.

3.2 Outputs

Part called: Outputs enable to users to view and to print the calculated values of the economic parameters, mainly: Chapter of the costs and benefits in financial numbers, economic results better say economic effectiveness designed building repair technology for the pavement construction presented by economic criteria like: Payback Period, Net Present Value, Internal Rate of Return. The user can view and print the general report for each action better say for all evaluated road sections and all evaluated building repair technology variants. Next the user can view and print the sequences of the urgency for the evaluated road section based on economic criteria values better say based on economic effectiveness of the investment to the road section repairing.
The mentioned outputs are sufficient and objective criteria for the program user (road administrator) for decision if the designed repair technology is effective, if this technology will be used for evaluated section or not. At the end road administrator can make decision if the evaluated road section will be repaired or not and decision making process will be correct and based on objective criteria.

### 3.3 Archiving

All data entered for each action can be archived. Then user can anytime use the data again, work with them, change them and use again. Archive in program serve in the same time like a tool for making of the sequences of the urgency based on economic criteria (Internal Rate of Return).

### 4. Final Decision Making Processes

The economic effectiveness expressed by economic criteria like IRR, NPV… are objective decision criteria but sometimes it is not enough. Stakeholders have to take into account all circumstances in this field of investment. There is different view for the optimal decision making processes and for the objective criteria. There are two various – opposite views. The first one is the view of road users as a road network customers and the second one is the view of the stakeholders (state government, road administrator leaders) as a budget “bodyguards”.

### 4.1 Definition of Requirements and Expectations of Customers and Other Stakeholders

The quality of construction work is defined by requirements specified by the state, the customer (final user) and the contractor. Required properties are defined in the relevant technical standards or are determined by the contractor in technical specifications.

In the past, the construction work was assessed only in terms of safety. This was not quite right. Today, more comprehensive view is preferred. If construction work is to meet the required parameters it must be acceptable in terms of:

- Reliability (safety, durability and usability)
- Quality Management
- Development and environmental protection [6].
4.2 **Defining the outcomes of the final process - the main features and objectives of the quality of roads**

Quality is a set of technical parameters that characterize the road within the project and enable it to carry out expected functions at the estimated time and under given conditions. Quality characteristics of "segments" - the functional nodes of the road network (during use) represent outputs of the final (top) processes of maintenance and renewal of the road network [9].

The quality of the road depends on unchanging geometrical and construction parameters of the road and also on variable transport and operation parameters.

The key quality criteria for the road are:

- Safety
- Cruising speed

These basic attributes of quality are essential factors and indicators of satisfaction of end users of the road.

Evaluation of quality and condition of the road is largely defined by the technical level of the road. This level is evaluated particularly by the following indicators:

- Transport, construction and technical condition
- Impact on the environment and regional development

From methods to measure performance of competition is in quality management systems in particular recommended the benchmarking. It is most often characterized as a systematic process of measuring and comparing the performance of products, services, processes and functions of organizations recognized as the basis for comparison in order to define their own improvement goals.

4.3 **Proposal of objectification of available resources allocation for the road maintenance and repair**

The subject is to propose optimization of the cost intensity of maintenance and repair of roads assigned to individual road administrators as an important tool promoting the efficiency of processes of roads maintenance and repair [13].

Analysis of indicators of efficiency of road maintenance and repair processes with the use of benchmarking detected significant differences in cost intensity at individual road administrators. This sought to emphasize specific conditions of administrators in quantifying the cost intensity of road maintenance and repair.

Enhancing the objectification of available funds allocation to individual road administrators is proposed by supplementing the impact of specific conditions of individual administrators into
the current method of funds allocation, as the decisive characteristics of road administrators. The impact of specific conditions is expressed by the impact of traffic load, technical conditions, climate conditions, geographical conditions and road width arrangements. Such approach will significantly strengthen the efficiency of costs used to repair and maintain the assigned portion of the road network, by increasing the degree of objectification at funds allocation. By integrating the above impacts into the method of allocating the resources, including the impact of traffic load, greater emphasis is put on the final road network users in terms of market economy.

4.4 Analysis and Selection of Potential Impacts on Cost Intensity

When examining the costs in previous periods at the level of basic processes, as already stated, significant differences were identified in cost intensity at individual administrators per kilometre of assigned roads. Analysis of reasons for different unit costs at various road administrators pointed out significant impact of especially five following factors which influence the cost of road maintenance and repair per kilometre the most.

- Traffic load - expressed in terms of total intensity or total transport capacity and heavy goods vehicles transport capacity
- Pavement technical condition - expressed mainly by the level of technical condition of roads on the basis of surveys conducted by the road administrator and residual life of roads
- Geographical conditions - expressed in particular by an average longitudinal inclination of roads
- Climate conditions - expressed in particular by an average number of days with significantly adverse weather (icy, freezing, precipitation, snow days)
- Width arrangements - expressed as an average road width per administrator.

5 Conclusions

Final decision making processes on the level of road administration should be depended on the objective criteria. The first and most important group of criteria is technical condition of the evaluated road section. The technical condition is expressed by non-variable parameters (they don’t change during the lifetime) and variable parameters. In Slovakia we have expert system (Pavement management system) for evaluation and prediction of these parameters of the pavement construction based on traffic volume and quality of the pavement construction. The goal of the presented system is to keep quality of the road network at least at the same level. The optimal scenario is to improve the road network quality level. So, the road
administrator has to allocate the resources with the using of the expert systems. One of them ISEH presented in this article is very useful tool for the decision making process. ISEH evaluates economic effectiveness of the investment to road maintenance and repair works by the economic criteria like Net Present Value, Internal Rate of Return and Payback Period. Economic criteria give the road administrator argument for decision to invest in evaluated road section or vice versa. Last few years experiences and our analyses show that economic effectiveness is objective criteria but if we need to achieve the basic goal - satisfaction of road users, we have to take into account comprehensive criteria for the road section quality. These criteria are: negative impacts from the traffic for the environment and inhabitants (emissions, noise, vibrations, dust…), attractiveness of the area where is the road section situated (often expressed by traffic volume), climate conditions (difference between north of Slovakia and south of Slovakia is for example about 5 degrees of Celsius or 25 cm snow cover during the winter). The conclusion is that we have developed quit good expert system for the road management in Slovakia, but it is necessary to continue in improving. There are a lot of challenges for Research and Development activities in the field of road management and we hope, we will present the next outputs of our activities in very short time.

This contribution is the result of the project implementation: "Independent Research of Civil Engineering Construction for Increase in Construction Elements Effectiveness" (ITMS: 26220220112) supported by the Research & Development Operational Programme funded by the ERDF.

This contribution is the result of the project implementation: „Supporting research and development center of excellence for civil engineering” (ITMS: 26220120031) supported by the Research & Development Operational Programme funded by the ERDF

„This work was supported by the Slovak Research and Development Agency under the contract No. LPP-0402-09“.

6 REFERENCES

Relationship between road width and safety

1 INTRODUCTION

Road accidents have made a serious contribution to death rate in the Czech Republic for a long time. In the course of time accident frequency trends changed several times, however current numbers are still unsatisfactory. Since 2002 there have been activities following the White paper (European Commission, 2001) with goal of halving the number of persons killed in road accidents by 2010 compared with 2002. A significant road safety improvement was achieved, however the goal was not reached. This situation calls for further strengthening of effort.

Safety, as a multidisciplinary component of the transportation system, is built on three pillars: the vehicle, the road user and the road environment (ie. geometry and other characteristics). According to PIARC (2003), an accident is considered a disruption of the balance among three mentioned pillars. The number of accidents is used to derive the measure of level of safety; typical variable, used in in the Czech Republic most often, is accident rate. However, number of studies (eg. Hauer, 1997) recommends using accident frequency instead. It is defined as a number of accidents in specific period of time.

2 ROAD WIDTH

Recalling mentioned pillars, this paper focuses on road geometry, and more precisely on road width. According to Gatti et al. (2007), it is the most appropriate value to characterize the cross section.

What is the effect of road width on accident frequency? According to Elvik et al. (2009), increasing the road width reduces the accident frequency on rural roads, but may lead to a small increase in accident frequency in urban areas. The reason is that road width in towns and cities makes crossings wider, so that pedestrians need more time to cross the road. In rural areas, increased road width may provide a safety margin, because speed is higher than in towns.
However, the road width consists of lane width and shoulder width, in some cases even the width of median barrier. All these parts make up the overall road width. Regarding relationship between shoulder width and safety, Elvik et al. (2009) conclude that wider shoulders almost always result in fewer accidents. Hauer (2000) states that shoulder width is more beneficial to safety at higher traffic volumes than at lower ones. Strathman et al. (2001) found more accidents on motorways with wider shoulders than on motorways with narrow shoulders. Gatti et al. (2007) lists both positive and negative aspects as well. In addition, all the results are not definite: Elvik et al. (2009) state that there is no relationship between road width and accidents when controlling for traffic volumes and speed (Garber and Erhard, 2000).

However, most of these results come from studies performed in Western Europe, Nordic countries and Northern America. These are all countries with certain level of road safety culture. It is questionable to think that the same will automatically hold in the Czech Republic as well. The paper searches for the answer to this question.

3 DATA

Data needed for road environment studies come from three data sets: accident data, road network data and traffic volume data.

Accident data
Accident data are reported by Czech Traffic Police. They include information on the accident, its location and all the vehicles and users involved. In total 59 items are registered for each accident. According to their consequences, accidents are sorted as property damage only (PDO) and injury accidents (including accidents with light injuries, serious injuries and fatalities). However with PDO accidents, there is a financial limit for reporting to the Police: while up to 2009 it was 50 000 CZK (approx. 2000 €), since 2009 it is 100 000 CZK (approx. 4000 €). Only accidents exceeding this limit are reported to the Police. Thus accident data suffer from underreporting.

Since 2007 all registered accidents are located with use of GPS which improves the location precision and further processing. In the Czech Republic three-year period is considered sufficient to base accident studies on – thus three complete years (2007, 2008 and 2009) of records were used. 2010 data were not yet available at the moment of writing.

Road network data
Road and traffic data were obtained from Road and Motorway Directorate of the Czech Republic (ŘSD). Road network data cover most of the Czech road network. They are
collected continuously and updated twice a year. In the paper data set from the first half-year 2007 was used in order to match with GPS located accident data starting since 2007. The part which is the most relevant to the topic is the *road passport*. It consists of interrelated databases of sections and intersections. Only sections were considered in the paper. Their database includes data related to each road section – road and lane width, presence of trees, description of median barrier, road surface, passive safety systems and many others. Section is defined as a segment of the road with uniform parameters. Whenever any parameter changes, the section ends and another one begins. Due to this fact, lengths of sections are not equal. In order to be able to use length in further calculation, section length values should have the same statistical distribution for both compared groups. The histograms of section lengths for both road width ranges are given in Fig. 1.

![Fig. 1 Histograms of section lengths for both road width ranges](image)

The two histograms of section lengths show similar shapes. To compare the distributions of both populations, a two-sample Kolmogorov-Smirnov test was performed. The test statistic is $D = 0.013$. The null hypothesis of different distributions was therefore rejected.
Traffic volume data
Traffic volume data come from National Traffic Census which is organized every five years. The last census was performed in 2010, however its results have not been published yet. Therefore data from 2005 were used.
Census data cover all the motorways, expressways, 1st and 2nd class roads. Some of 3rd class roads are not included. The results of traffic census include annual average daily traffic volume (AADT) on every section in 13 vehicle categories.

Data selection
As mentioned above, data sets provided by ŘSD were very large. In order to form relatively homogeneous group, selection of road width category had to be made. Therefore 9.5m and 11.5m roads were selected. According to České dálnice (2010), these categories of roads are the most frequent in the Czech road network (excluding expressways and motorways). Only two-lane road sections (with one lane in each direction) were selected.
These categories fit to the Czech national standard for design of highways and motorways (ÚNMZ, 2009). Design categories parameters are listed in Tab. 1.

Tab. 1 Design parameters of selected road categories according to ÚNMZ (2009)

<table>
<thead>
<tr>
<th>design category</th>
<th>width</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b [m]</td>
</tr>
<tr>
<td>9.5</td>
<td>3.50</td>
</tr>
<tr>
<td>11.5</td>
<td>3.50</td>
</tr>
</tbody>
</table>

The following symbols are used:

\( b \) … road width
\( a \) … lane width
\( v \) … width of carriageway marking
\( c \) … width of paved shoulder
\( e \) … width of unpaved shoulder

Values in Tab. 1 show that the only difference between 9.5m and 11.5m roads is the shoulder width. The meaning of symbols is illustrated in Fig. 2.
4 METHODS

The aim is to analyze road width and its impact on safety. The data selection includes two-lane roads of 9.5m and 11.5m width (single carriageway with no median barrier). Selection was not random – it covered the whole Czech Republic. Following steps – including stepwise reduction of selected data set – were performed:

4.1 Rejection of PDO accidents
Due to mentioned legislative change of reporting limit, number of registered PDO accidents fell down significantly in 2009, biasing the statistics. On the contrary, the numbers of injury accidents (light injuries, serious injuries and fatalities) are not biased by this change – see Fig. 3.

In the paper, period between 2007 and 2009 is studied. To overcome the mentioned bias, only injury accidents were used in the calculation.
4.2 Rejection of non related causes

General accident causation comes from Police accident investigation. While it is questionable and may be subject of thorough analysis, the Police only assigns one single cause to each accident, chosen in a pre-defined list. However not all the causes are related to the purpose of this study – for example accidents due to failure to yield the right-of-way or vehicle technical failure. Following the above mentioned statement by Elvik et al., 2009 (*the increase of width should provide more space for overtaking and speeding, resulting in lower accident frequency*), the accidents considered were restricted to overtaking and speeding causes, which induce the most serious accidents. The same restriction was applied also by Kafoňková and Andres (2008). The final list included thirteen causes (see Tab. 2).

<table>
<thead>
<tr>
<th>Tab. 2 Selected causes of accidents for this study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speeding</strong></td>
</tr>
<tr>
<td>- maladjustment of speed to traffic operation density</td>
</tr>
<tr>
<td>- speed higher than speed according to traffic rules</td>
</tr>
<tr>
<td>- speed higher than speed indicated by traffic sign</td>
</tr>
<tr>
<td><strong>Incorrect overtaking</strong></td>
</tr>
<tr>
<td>- overtaking to the right</td>
</tr>
<tr>
<td>- overtaking without sufficient side distance</td>
</tr>
<tr>
<td>- overtaking without sufficient sight (at the curve or its proximity, in front of summit etc.)</td>
</tr>
<tr>
<td>- during overtaking a driver in opposite direction was endangered (wrong estimate of distance etc.)</td>
</tr>
<tr>
<td>- during overtaking an overtaken driver was endangered (enforced entering, violent braking, change of ride direction etc.)</td>
</tr>
<tr>
<td>- overtaking in places where the overtaking is forbidden by traffic sign</td>
</tr>
<tr>
<td>- overstepping the separation line during overtaking</td>
</tr>
<tr>
<td>- overlooking the vehicle already overtaking</td>
</tr>
<tr>
<td>- other kind of incorrect overtaking</td>
</tr>
<tr>
<td><strong>Incorrect driving</strong></td>
</tr>
<tr>
<td>- driving on the wrong side, entering the opposite direction</td>
</tr>
</tbody>
</table>

4.3 Grouping according to traffic conditions

To make valid safety comparisons, Hauer (2005) recommends to compare the expected accident frequencies under the same conditions. Such conditions include traffic speed and traffic volume.

Speed limits in the Czech Republic are set to 50 kph in urban areas and 90 kph in rural areas. It is therefore important to distinguish between sections in urban and rural areas. Such grouping is commonly applied in safety performance functions – see eg. Reurings et al. (2005). This report also states that traffic volume is the most significant factor for accident frequency. Traffic volume conditions were defined so that they reflect the Czech national standard (ÚNMZ, 2009) design volume limits. The limits are 10 000 vpd on 9.5m roads and 12 000 vpd on 11.5m roads. Therefore traffic volume conditions were set as follows:
– according to the design limit (up to 10 000 vpd on 9.5m roads, up to 12 000 vpd on 11.5m roads)
– exceeding the design limit (over 10 000 vpd on 9.5m roads, over 12 000 vpd on 11.5m roads)

These two volume states were named as *standard* and *high*; their sum is *total*.

In order to take volume into consideration, only sections with known AADT were used for the study. As it was mentioned in chapter 3, the census results cover most of the Czech road network. Because they reflect the 2005 census, AADT values were corrected to values of 2007 – 2009; for example accident in 2008 was linked with the expected volume in 2008. The sections with no accidents were assigned the volume in 2007. This correction was made according to the Czech technical guidelines TP 225 (Bartoš et al., 2010).

### 4.4 Calculation of accident frequency

To quantify the relationship between road width and safety, accident frequency (AF) was used. It is defined as follows (see eg. Hauer et al., 2002):

\[
AF = \frac{N}{L \cdot Y}
\]

where \(N\) is number of injury accidents on the section, \(L\) is length of section [km], \(Y\) is duration [year].

To illustrate the calculation, if 2 injury accidents happen at 500m section between 2007 and 2009 (3 years), the accident frequency is:

\[
AF = \frac{2}{0.5 \cdot 3} \approx 1.3
\]

It means that approximately 1.3 injury accidents per 1 km and 1 year may be expected at similar section, ie. with similar traffic conditions such as speed limit or traffic volume.

### 5 RESULTS

As it was mentioned in chapter 4.3, road sections were grouped according to traffic conditions – see Tab. 3. For each row in the table null hypothesis was assumed. For example null hypothesis 1 assumes that distribution of AF on 9.5m roads in urban areas for total volume is equal to distribution of AF on 11.5m roads in urban areas for total volume.
Tab. 3 List of groups and numbers of assigned null hypotheses

<table>
<thead>
<tr>
<th>number of null hypothesis</th>
<th>group 1</th>
<th></th>
<th>group 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>width</td>
<td>area</td>
<td>volume</td>
<td>width</td>
</tr>
<tr>
<td>1</td>
<td>9.5 m</td>
<td>urban</td>
<td>total</td>
<td>11.5 m</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>standard</td>
<td>standard</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>high</td>
<td>high</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>9.5 m</td>
<td>total</td>
<td></td>
<td>11.5 m</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>high</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The statistical standard is to use Student’s t-test to compare variable means which is however applicable for normal distribution only. The number of accidents and their frequency are typically considered to have negative binomial distribution (see e.g. Lord and Mannering, 2010). Therefore Mann-Whitney U test was chosen, being an alternative to Student’s t-test applicable for other distributions. It was used to test difference between distributions of group 1 (9.5m road sections) and group 2 (11.5m road sections). Null hypothesis assumed that both group have equal distributions. Results are listed in Tab. 4.

Tab. 4 Values of accident frequency and related test characteristics

<table>
<thead>
<tr>
<th>number of null hypothesis</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample size</td>
<td>1921</td>
<td>809</td>
<td>1435</td>
<td>478</td>
<td>487</td>
<td>331</td>
</tr>
<tr>
<td>AF mean</td>
<td>0.08</td>
<td>0.18</td>
<td>0.02</td>
<td>0.20</td>
<td>0.27</td>
<td>0.16</td>
</tr>
<tr>
<td>AF variance</td>
<td>0.94</td>
<td>5.49</td>
<td>0.07</td>
<td>8.11</td>
<td>3.52</td>
<td>1.72</td>
</tr>
<tr>
<td>significance</td>
<td>0.272</td>
<td>0.081</td>
<td>0.376</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>result</td>
<td>null hypothesis retained</td>
<td>null hypothesis retained</td>
<td>null hypothesis retained</td>
<td>null hypothesis rejected</td>
<td>null hypothesis retained</td>
<td>null hypothesis retained</td>
</tr>
</tbody>
</table>

All but one hypothesis were retained – their differences are not significant. The only significant difference is under the fourth hypothesis, i.e. between 9.5m and 11.5m rural roads at total traffic volume. The significance of sixth hypothesis is 0.051 which is practically almost at the level of null hypothesis rejection ($p = 0.05$) as well.
6 CONCLUSIONS AND DISCUSSION

According to results of statistical tests it can be stated that there is no difference between AF on urban roads (null hypotheses 1, 2, 3 were retained); however there is a difference between AF of 9.5m and 11.5m rural roads at total and high volume (null hypotheses 4 and 6 were rejected).

The observed difference has two aspects:

1) At total volume the difference between AF means on 9.5m and 11.5m rural roads is almost twofold (0.07 vs 0.13).

2) Results are opposite at rural sections with high volume only: their AF at 9.5m sections is higher that at 11.5m sections (0.21 vs 0.13).

The reason of difference (1) may lie in insufficient safety culture in the Czech Republic, which causes more speeding and overtaking related accidents on wider roads.

The result (2) falls in line with finding of Hauer (2000) who states that shoulder width is more beneficial to safety at higher traffic volumes than at lower ones. He sees the whole relation as sum of a number of conflicting tendencies, including beneficial recovery area provision. This is also mentioned provided by Stein and Neuman (2007) among the safety functions of shoulder: shoulders may increase safety by providing a recovery area for drivers who have left the travel lane while attempting to avoid a crash or an object in the lane ahead. All these opposite tendencies may cause the above mentioned inconsistency.

To sum up, relation between road width and safety (in terms of accident frequency) was found to differ significantly on Czech 9.5m and 11.5m rural roads. It may be attributed to combination of speeding and overtaking (more frequent at 11.5m roads) and being provided less recovery space (typical at 9.5m roads) at higher traffic volumes. These facts are in fact complementary and overlapping. It can be concluded that drivers behave the same way, regardless on the road width, however they have less space for overtaking and recovery manoeuvres at 9.5m roads.

All these assumptions need to be checked and assessed by further research. Future studies might also consider other traffic conditions to set the categories to be tested. The ultimate solution would be developing accident prediction model, taking into account more variables in order to describe the traffic conditions in more complex way. All these data are contained in road passport introduced in chapter 3. Such a model should use equal length of sections and predict expected number of accidents. It could also differ between accident severity, type of road (motorways, expressways…) or geometric variables of the road.
REFERENCES


ÚNMZ (2009). ČSN 73 6101 *Projektování silnic a dálnic*. 

Transport, Civil and Road Engineering
EMERGENCY VEHICLE PRIORITIZATION USING VEHICLE-TO-INFRASTRUCTURE COMMUNICATION

Keywords: V2I communication, Emergency vehicle, traffic management, traffic simulation
Research domain: Intelligent Transport Systems (ITS)
ABSTRACT

Emergency vehicles need to reach their destination as fast as possible. They deserve the highest priority at intersections. Therefore, they are allowed to use bus lanes and pass red lights at traffic light systems. Nevertheless, for emergency vehicles it is still quicker and safer to get priority at the approaching intersection. This paper analyses how the travel time of emergency vehicles can be improved by using vehicle-to-infrastructure (V2I) communication. Emergency vehicles are sending messages with their route information and their current position. The traffic lights which have to be passed can switch to green for emergency vehicles and to red for all other streets. The traffic lights continue the normal operation after the emergency vehicle has passed the intersection. Simulation results show that emergency vehicles can reach their destination faster.

1. INTRODUCTION

This section describes the traffic problem which is addressed and the main objectives of this research work.

1.1 CONTEXT AND PROBLEM DEFINITION

Emergency vehicles can be divided into three classes: ambulances, fire trucks and police cars (Becker, 2002). In case of an emergency these vehicles have special rights e.g. they are allowed to violate red lights, use bus lanes and overtake other vehicles. Emergency vehicles are only allowed to use these special rights if it is necessary to reach the destination as fast as possible to save human lives or prevent serious injuries (Hempel, 2007). A blue flashing light and a siren is signalling an approaching emergency vehicle and requests priority at intersections and the right to overtake. However, emergency vehicles are not allowed to force other traffic participants to give them priority. Some drivers do not realize that an emergency vehicle is approaching because they are listening to loud music or they are absent minded. Major accidents might happen if an emergency vehicle expects to have priority but other drivers do not realize that an emergency vehicle is approaching. Consequently, a driver of an emergency vehicle has to stop at a red light and has to decide whether it is possible to violate the red light without endangering other people. Statistically, emergency vehicles are more likely to be involved in car accidents than other vehicles (Hempel, 2007). Driving in case of an emergency is a stressful situation. The driver is in conflict between reaching the destination as fast as possible and driving slower but carefully to avoid accidents. It would be safer and
faster if emergency vehicles get priority at every controlled intersection by the traffic light system.

1.2 OBJECTIVES OF THE PAPER

The objective of this work is to develop a strategy to improve the routes of emergency vehicles regarding travel time and safety. Therefore, a simulation model for emergency vehicles has to be implemented in microscopic traffic simulation SUMO. Furthermore, traffic management strategies for emergency vehicles using cooperative communication technologies will be analyzed. The aim is to answer the following questions: how can the travel time of emergency vehicles be improved? How can dangerous situations e.g. violating red traffic lights be avoided?

The paper is structured as following: at first, a brief introduction of special rights and the risks of emergency vehicles are given. Then, Vehicle-to-Vehicle and Vehicle-to-Infrastructure communication is described. Next, the used simulation scenario is described and the results of the simulation are given. The document ends with a discussion.

2. STATE-OF-THE-ART

This section gives a brief overview of existing literature of V2X communication and this research field contributes to this research work.

2.1 LITERATURE AND RESEARCH REVIEW

The potential of V2X communication is large. On the one hand, V2X can help to make driving safer. The driver could be warned if another vehicle is approaching the same intersection or if there is a construction site or if the street is slippery because of lost oil etc. On the other side, V2X can improve traffic management. RSUs can collect the travel information of equipped vehicles. This information is helpful for other vehicles to decide which route is faster depending on the current traffic state. Using V2X communication, a driver can be informed about a traffic jam and the recommendation to use another faster route instead. In this research, the IEEE 802.11p standard is used for the message transmission of the V2X communication. In this research, only Cooperative Awareness Messages (CAMs) are considered. CAMs are periodically sent and include current information of the sender e.g. the position of the vehicle.
V2X communication might help preventing accidents of emergency vehicles or help them to reach their destination faster. One idea is that emergency vehicles are sending an “emergency vehicle approaching warning” (Strang, 2010). All vehicles which are able to receive such messages will be notified about the approaching emergency vehicle and can react accordingly.

The scope of this research is not only on informing drivers about an approaching emergency vehicle but rather on giving emergency vehicles priority at controlled intersection. Nelson (2007) already evaluated the impacts of emergency vehicle priority systems. A traffic controller changes manually a traffic light to give priority to emergency vehicles or to other traffic participants. The idea was to change the duration of the green phases to free the lanes before an emergency vehicle is reaching the traffic light. Many people think that this research approach would cause significant disruptions in the traffic flow, but Nelson (2007) showed in his study that this is not always the case.

2.2 CONTRIBUTION

The idea in this research is to give priority to emergency vehicles at controlled intersection via V2I communication. The main difference to the study of Nelson (2007) is that this approach is using V2I communication and is working automatically, while Nelson (2007) needed a traffic controller who manipulates the traffic lights manually. The traffic lights are switched to green for emergency vehicles while all other vehicles have to wait at a red light. The procedure is as following: An emergency vehicle on its way to the hospital or to an incident location sends unicast messages to a road side unit, which include its expected route and current position. The traffic management center is informed about an approaching emergency vehicle when a road side unit (RSU) receives such a message. The traffic management center sets the traffic light in front of the oncoming emergency vehicle to green while all other vehicles have to stop at a red light. The traffic light controller can continue its normal operation after the emergency vehicle has passed the intersection. This procedure is repeated for every controlled intersection on the route of the emergency vehicle.

3. SIMULATION AND RESULTS

A lot of research effort is put into the development of communication between vehicles and infrastructure (V2X), but real world tests are expensive, time consuming and might even be dangerous. Therefore the interest in simulating the effects of V2X applications before implementing them in the real world rises. The implementation of an open-source system which allows such simulations was addressed by the project “iTETRIS” (iTETRIS, 2011). It was performed between July 2008 and January 2011 and was co-funded by the European Commission. Two well-known simulators were connected within a middleware called
“iTETRIS Control System” (iCS) (Gozalvez et. al., 2009; Lazaro, et. al., 2008). The simulator ns3 (ns3, 2011) was used for the V2X communication and SUMO (Krajzewicz et. al., 2002; SUMO, 2011) for the road traffic simulation. The simulators and other additional programs which are simulating the V2X application are started by the iCS. Additionally, the main task of the iCS is to synchronize the simulators and the exchange of messages between the programs.

3.1 SCENARIO DESCRIPTION

For this study the traffic of the city of Bologna was simulated with iTETRIS. Before setting up the Bologna scenario, the needs and problems of Bologna were evaluated. For the evaluation the traffic data of Bologna was collected and analysed. Afterwards possible ways for improving the traffic conditions were determined. The streets of Bologna are mostly small and narrow which causes many traffic jams. In case of an emergency it is difficult for drivers of emergency vehicles to pass these crowded streets. Several other problematic areas and ideas on V2X applications to solve these problems were investigated in the “iTETRIS” project (see Blokpoel et. al., 2010). iTETRIS concentrated only on traffic management applications, in contrast to other projects co-founded by the European Commission. As a matter of fact, the focus of this research is on traffic management strategies and not on safety or infotainment related applications.

Figure 1: Region of Bologna which was used for the simulation (Cartolano, et. al., 2010)

For the simulation the collected traffic data and a set of network scenarios with different sizes were given by the administration of the city of Bologna which was one of the partners of
iTETRIS. The network information was supplied as Vissim (PTV, 2011a) and VISUM (PTV, 2011b) input files.

One of the scenario networks was the region pictured in Figure 1. The scenario was chosen because there are many traffic jams. Furthermore, there is a football stadium within the region. This leads to an additional traffic flow which also causes large traffic jams. The modelled emergency vehicles are allowed to use all roads, including those reserved for buses, the network contains also such lanes. Figure 3 shows the simulated route of the emergency vehicle.

The road network is equipped with RSU at traffic light controlled intersections. The emergency vehicle sends CAMs which identify it as a higher prioritized emergency vehicle. As soon as such a CAM is received by an RSU, the RSU triggers the assigned traffic light to turn green for the direction the emergency vehicle comes from. The direction is determined using the vehicle’s lane information stored in the CAM. After the emergency vehicle has passed the intersection, determined again using the lane information from a subsequent CAM, the traffic lights are set to their original program. We simulate a single emergency vehicle which runs the route shown in Figure 2 and starts at second 1800. The original demand of the "joined" scenario which represents the peak hour between 8:00 and 9:00 is used. We run the simulation for half an hour, assuring the traffic is in a stable state. After this time, the emergency vehicle starts its ride. Only the simulated emergency vehicle is interacting with the RSUs; as the RSUs must only be informed once about the emergency vehicle's arrival. In this research the possible loss of messages due to vehicle interaction was not considered.

3.2 SIMULATION

It should be noted that to our knowledge, no proper model for simulating an emergency vehicle exists in the literature. While running the scenario with the application, and the plain
vehicle movement model, we encountered the following situation: as vehicles normally have to wait at roundabouts and our emergency vehicle has to pass one, it got stuck when approaching it - waiting for almost a minute for a free slot. This of course does not resemble the reality. In probably most western countries, normal vehicles would wait to let the emergency vehicle enter and cross the roundabout. Accordingly, the simulation was extended and performed regarding this aspect. Figure 3 shows a roundabout of the simulation scenario. The vehicles which are driving inside of the roundabout normally have the right of way. But when an emergency vehicle is approaching they should give way to the emergency vehicle. For simulating this behaviour all vehicles which are driving on a street inside of the roundabout are stopping if an emergency vehicle is reaching the roundabout.

Other interactions with an emergency vehicle, such as freeing a lane to let an emergency vehicle pass or cross an intersection as first, or assuring an emergency vehicle can cross an intersection at red lights, etc. must be developed in further steps.

![Figure 3: The traffic light signal switches to green for the emergency vehicle (left). Cars give priority to an emergency vehicle at a roundabout (right).](image)

For the simulations done here, we did not encounter any further major drawbacks after having implemented the higher priority at roundabouts, mainly because the emergency vehicle runs over a green light district in any case, by accident. Resulting, only one traffic light is influenced by the application. The following results are in fact even more impressive regarding this fact.

### 3.3 Evaluation and Results

At first, we look at the emergency vehicle's performance. Figure 4 compares the travel time and the waiting time of the ride between both, a plain simulation of the emergency vehicle's ride and for the one equipped with the prioritization application.
As one can see, the equipped emergency vehicle reaches its destination 40 s earlier than a non-equipped one, and the waiting time is reduced. The emission values were found to be almost similar, with a slightly worse result for CO emissions for the equipped emergency vehicle, which is assumed to be able to be neglected. The 40 s were gained at a single intersection, as the non-equipped emergency vehicle was waiting at the end of a queue which needed two cycles to let the emergency vehicle cross the intersection. When running the application, all waiting vehicles could drive over the intersection within the prolonged cycle. For the chosen route, this was the only intersection at which the emergency vehicle had to wait. The other controlled intersections passed by the emergency vehicle are located at bus lanes, and due to this are not opposing queue formation in their front. As complete streets dedicated to busses only are relatively seldom, one could expect a large benefit for all passed intersections.

Besides prising the benefits for the emergency vehicle, the effects on the other traffic participants should be investigated, too. When looking at the whole simulated vehicle population, the drawbacks for the normal participants seem to be minor, as shown in Figure 5.

But not all vehicles are affected in the same way. Vehicles crossing the intersection in the same direction as the emergency vehicle gain from having a longer green time while vehicles
crossing the intersection in perpendicular direction have to wait longer and probably earlier. In our case, obviously more vehicles benefit.

The simulated application is very brute-force. It should be revalidated, taking into account possible legal rules about switching phases more explicitly. This may yield into the reduction of the travel time benefit by some seconds for the emergency vehicle. There are vehicles which definitely have to wait longer on intersections to let the emergency vehicles pass, and the application's performance can probably be improved. As an example, the distance in front of the emergency vehicle at which the traffic lights switches to green could be optimized.

4. CONCLUSIONS

This section provides a conclusion of the research work in this paper and gives recommendations for further research works.

4.1 CONCLUDING REMARKS

The simulation shows that V2I communication can help to improve the travel time of emergency vehicles. The safety aspect could not be simulated within this simulation. In the simulation SUMO it is not possible to simulate accidents. Furthermore the simulated vehicles are always following speed limits and traffic rules. Consequently, it is not yet possible to simulate whether safety could be improved because of the V2I strategy. The application seems to be very promising, as saving some seconds per intersection can save a human life.

Another advantage of this application is than there is no need for a high penetration rate of equipped vehicles. Many cooperative applications have the problem that there is only a benefit if a significant percentage of vehicles are equipped. Consequently, the early adapters of V2X communication will have to spend a lot of money without having a real benefit.

4.2 RECOMMENDATIONS FOR FURTHER RESEARCH

In real life an emergency vehicle would probably be faster than the simulated vehicles because other vehicles let the emergency vehicle pass and the emergency vehicle is able to violate red lights. The simulation of a realistic behaviour of an emergency vehicle should be implemented and analysed in further works. Other not yet implemented aspects are to exceed the maximum speed of a street and to use a street in the wrong direction to overtake other vehicle.
In a further step it should be investigated whether the routing of emergency vehicles could be improved by using V2I and V2V communication. By using V2X communication the current traffic state of the streets can be interfered. Emergency vehicles could use this information to decide which route is the fastest route depending on the current traffic state.

Another interesting research question is how drivers react to the traffic management strategies. For example what will happen when the traffic light is red for a long time but they can not see an emergency vehicle? Or another aspect is how people react when they receive a message about an approaching vehicle. Will they slow down, take another route or will they ignore the message?

REFERENCES


THE IMPACT OF TRAFFIC MANAGEMENT OPERATIONS ON MOTORWAY CONGESTION:
A LEVEL OF SERVICE BASED EVALUATION METHOD
ABSTRACT

Existing methods for assessing the impact of traffic management operations on congestion use dynamic traffic simulation tools for ex-ante evaluation, while ex-post evaluation is carried out by before/after comparisons. However, simulation of the interactions among the different stretches within a given network requires a significant amount of data that are not always available. In addition, before/after comparisons cannot single out the real impact of a traffic management operation from that of a possible demand change over the network. We propose a method that addresses these two constraints using link-aggregated traffic data obtained from loop detectors. Those data are the total vehicle kilometres travelled (VHT), the total vehicle hours travelled (VHT) and the travel time (TT). We show that it is possible to retrieve the demand matrix corresponding to any observed traffic conditions on a motorway stretch. This matrix may then be modified to represent a traffic growth or diversion. Furthermore, link-aggregated data also provide the capacity of the stretch, which in turn may also be modified to represent the effect of some traffic management operations, such as lane management. The method is validated with data from the A1 motorway in France, for which two application examples are also given.

1. INTRODUCTION

Traffic management aims at tackling growing congestion on motorway networks. Depending on sites and traffic characteristics, the main variants are managed lanes, variable speed limits, ramp metering and traffic routing. Their action consists of improving traffic conditions by increasing the capacity of the facility or by decreasing the demand rate at the bottlenecks. Most published articles related to the assessment of the impacts on congestion concern ex-post evaluations (Kellermann et al, 2000; Schrijnen, 2001; Shahin, 2003; Cohen, 2005; Cohen, 2007). Regarding lane management, those studies analyse changes in capacity and mean speeds. They usually conclude that such operations lead to capacity increases. All other things being equal, this increase is responsible for all changes in traffic conditions. As noted by Sultan et al (2008) in their evaluation report of the Active Traffic Management implementation in the United Kingdom, the effect of a capacity increase may be underestimated in case of a traffic growth. However, the authors do not propose any method to single out the only impact of the operation on the observed mean speeds, travel times or any other indicator related to congestion.

One method widely employed in a posteriori assessments of the impacts of traffic management operations consists in using control stretches. The central idea, as described by Hauer (1999), is that the control stretch is affected in the same manner by all other factors that affect the stretch under study, except by the traffic management operation. Several studies use this method either to test changes in drivers’ behaviour, or to analyse the safety impacts of an operation. For example, Rämä and Kulmala (2000) studied changes in mean speeds on three
sites after the installation of variable message signs (VMS) indicating the roads slipperiness and recommending minimum headways between vehicles during winter. The effect of the VMS on mean speeds was obtained by excluding changes observed on mean speeds at locations where the operation was not in force. While this method is especially appropriate for discrete variables, such as mean speeds or accident rates, it can hardly be applied to assess impacts on congestion. Traffic management operations are usually implemented during peak periods but their effects may also extend to off-peak periods. The first problem would be to find a control stretch on which traffic conditions would have the same daily pattern and that could be affected in the same way by a demand growth for instance. We know, from the shape of the fundamental diagram, that depending on the level of congestion the same demand growth rate may not affect two different stretches to the same extent. Furthermore, in this case the control stretch should also have the same origin and destination than the stretch under study, in order to ensure the same trips’ characteristics (especially for traffic composition).

Traffic simulation tools are usually not employed in a posteriori evaluations, essentially due to the volume of data required or available to recreate at this stage the reference situation. On the other hand, these tools are very useful in a priori evaluations where they facilitate the analysis of traffic growth for instance. Aparicio et al. (2004) took up a previous study of Monzón et al. (2000) related to a High Occupancy Vehicle (HOV) lane in Madrid. A discrete choice model is coupled with a macroscopic simulation model to estimate changes in traffic conditions on the adjacent general purpose lanes. Interactions with other stretches on the network are not considered.

Traffic diversion is mentioned in the evaluation study of the impact of the conversion of a general purpose lane into an HOV lane in Norway by Haugen (2004). Haugen noted that traffic volumes did not rise on parallel routes because the travel time increases were not significant enough on the remaining general-purpose lanes adjacent to the dedicated lane. In a study aiming at predicting the impact of HOV lanes on access ramps, Rodriguez et al (2008) analysed traffic diversion within the motorway corridor including the motorway stretch under study and parallel arterials. The main hypothesis was that drivers will divert from those arterials to the motorway if their travel time-saving is at least 5 minutes. Although this method allowed predicting the changes on the motorway congestion, the overall impact for the whole corridor was not estimated.

Depending on the country, several indicators are used to characterise congestion over a motorway or network. In France, the HKM (hour-kilometres) of congestion is currently one of the most used by traffic operators. It is the total length of congestion, brought on a single lane, multiplied by the duration of the observation (Mulot and Buisson, 2006). For this indicator, congestion is usually defined as traffic conditions with mean speeds at 30 km/h or lower, however a traffic operator may set another threshold. The HKM evolution may be analysed throughout the day and especially during peak hours. However, when considering an
impact evaluation, the main disadvantage of this indicator is that its relationship to traffic demand is not explicit. For instance, a reference document (CGPC, 2004) suggests a 2.9% growth of the HKM for a 1-3% increase in the total traffic demand, but does not consider the daily variations in traffic. Furthermore, regarding a socioeconomic evaluation, there is no value for the congested HKM, but rather of the time spent by travellers within this time-space. It is however possible to introduce the stretch density to convert the HKM into vehicle hours travelled (Cohen, 1999; Princeton, 2007). The latter monetisable indicator, widely employed, easily associated to the demand level, is the one we will use in this study to elaborate an evaluation method for the impacts of traffic management operations on congestion.

The approach proposed in this paper consists in analysing the evolution of traffic conditions at the link scale by using three variables: VKT, VHT and TT. They represent the total vehicle kilometres travelled, the total vehicle hours travelled and link travel time, respectively. We show that it is possible to retrieve from those variables the demand matrix of the link for a given period. This matrix may then be partly or totally modified to represent a traffic growth or diversion. The application of the approach is straightforward in ex-post evaluations, especially when analysing the impacts of a capacity increase or a demand-rate modulation. It may also be a relevant tool for large-scale studies while taking into account the interactions among stretches following the implementation of a traffic management operation, as well as their safety and environmental impacts.

The paper is organised as follows. Part 2 describes the datasets and tools used in this work. Part 3 presents observation results that highlight some existing relationships between the three aforementioned variables. Part 4 shows the application of these relationships into the proposed evaluation method. We finally conclude with some considerations and discussion on future steps.

2. METHOD

2.1. Site Description

The main site under study is a 17-km stretch of the A1 motorway between Charles de Gaulle Airport and Paris (southbound). It comprises 3 to 5-lane sections, six access ramps and six exit ramps. This motorway is one of the busiest in France, with daily traffic averaging 130,000 vehicles in both directions. Congestion (with criterion mean speed less than 30 km/h) lasts more than 5 hours during the AM peak period. Queue lengths usually exceed 5 kms, upstream from Paris. We also briefly present a few observation data for another urban motorway stretch near Paris, the A3.
2.2. Data

The A1 stretch is equipped with traffic flow, mean speed and occupancy rate measurement stations which are based on inductive loops. The distance between adjacent stations ranges from 500 m to 1 km. These are part of the SIRIUS system, which aims at optimising the traffic conditions on the motorway network of the Paris region through all sorts of traffic management operations. The SIRIUS database collects traffic data and, for each link, computes measures of performance such as the VKT, VHT and TT, which are used in this study as one-hour averages. One feature of SIRIUS is that at each time interval, the VKTs are computed for the whole stretch on the one hand, and the congested parts only, on the other hand. By dividing these two measures by the length of the stretch, we obtain the Generalised Volumes (GV) and Generalised Congested Volumes (GCV) respectively, as referred to in the rest of the paper. In like manner, a Generalised mean speed is also computed for the whole stretch. It is given by the ratio: \[
\frac{\text{VKT}}{\text{VHT}}.
\]

All data used in this study represent typical days during the period from April to June 2008 and September to November 2009. Typical days are working and class days with good weather conditions and no incidents or accidents. We also ensure that the daily reliability rate of each of the 20 monitoring stations that are installed all along the A1 stretch is 80% or higher. The reliability rate is defined as the ratio between the number of valid data and the total number of data expected over a given period.

2.3. Simulation Tool

The simulation tool used is FREQ12, a macroscopic model based on the relationship between traffic basic variables, the conservation equation of vehicles and the fundamental diagram. The simulator is tuned by direct manipulations of the section capacities (Gomes et al., 2004). As all simulation packages, FREQ12 has its own restrictions that the user needs to take into account during simulation. For instance, the model only allows a maximum of 24 time steps for simulation which greatly reduces the temporal limits of the study. In order to capture the daily traffic conditions we used data aggregated over 1-hour time intervals. Another restriction of FREQ12 pertains to the sections fundamental diagrams. The model provides a set of curves among which the user must choose the one that best fits the real data. The minimum value for free-flow speed is set at 50 mph (~80km/h).

3. OBSERVATION

3.1. Generalised Fundamental Diagram of a Motorway Link

Generalised Fundamental Diagram of a motorway link is a relatively new notion analysed by Cassidy (2011) who relies on former work (Edie, 1963; Geroliminis and Daganzo, 2007;
Buisson and Ladier, 2009) on the existence of a macroscopic fundamental diagram, MFD, describing traffic operation over a network in urban areas. In his study, Cassidy showed that a motorway link may be considered as a succession of sections with different characteristics (number and width of the lanes, slope, etc). This link has a macroscopic fundamental diagram which represents the outer envelope of all the (VKT, VHT) pairs of the whole link. Nevertheless Cassidy concludes by pointing out that those pairs are rarely reached since they correspond to traffic situations where traffic on all the sections would be under the same regime, congested or uncongested. On the other hand most of the (VKT, VHT) pairs fall under the MFD. In the light of this observation, in this study we will consider the mean fundamental diagram, which includes all observed pairs regardless of the traffic state on the different sections. This diagram will be called the Generalised Fundamental Diagram (GFD) of the link, as illustrated in the two graphs below (red dots in Figure 1) for two motorway stretches in the Paris region, A1 and A3.

Calibration of the fundamental diagrams of the different sections showed that the capacity given by the Generalised Fundamental Diagram corresponded to the bottleneck capacity, which was the smallest capacity along the whole stretch. It may serve to define the practical capacity used in link capacity functions (Branston, 1976) for static traffic assignment and which relate travel times to the number of vehicles. This question will be analysed in more details in section 2.3.

### 3.2. Vehicles Storage on a Link

Loop detectors installed along a stretch may give insight into the number of sections on which traffic is congested, as well as the number of vehicles concerned. One only needs to consider those sections with a mean speed lower than the critical speed of the Generalised Fundamental Diagram (about 60 km/h for both stretches in the previous graphs). The following graphs (Figure 2) depict the variation in the Generalised Congested Volumes (GCV) on the A1 motorway for two different days. The days were chosen because of their very different levels of congestion, about a 1 : 2 ratio for the morning period.
We noticed that GCVs appear in the morning by 6-7H, grow gradually until the peak hour and then decrease or disappear at mid-day. These volumes rise again as the evening peak approaches and dissipate by 21-22H. In this section we show that these GCVs represent the volume of vehicles that are stored on the link. The storage on a road section during a time interval is the difference between the demand for that period and the capacity of the section (May and Leiman, 2005). Since we already have the link capacity from the Generalised Fundamental Diagram, we will use a simple algorithm that allows us to retrieve the demand matrix corresponding to any given change of the GCVs (like the ones shown in Figure 2). The following algorithm is based on the conservation law for vehicles in a road section, and is used in traffic simulation tools.

The main notations are given below:

- $k = 1, 2, ..., N$ represent the time intervals of the study; traffic is assumed to be free-flowing at both $k=1$ and $k=N$.
- $D_k$: Traffic demand at time interval $k$;
- $Q_{k,x}$: Outflow at time interval $k$;
- $Q_k$: Total volume on the entire stretch at time interval $k$ (obtained from VKT for all sections);
- $Q_{k,sat}$: Congested volume at time interval $k$ (obtained from VKT for congested sections only);
- $Q_c$: Capacity of the stretch (obtained from the GFD);

0. Set $D_1 = Q_{k=1} = Q_{k=1,x}$ et $Q_{k=1,sat} = 0$, for the first time interval $k=1$, then proceed with steps 1 to 5 for time intervals $k=2$ à $k=N$.

1. Compute $Q_{k,x} = \begin{cases} Q_k & \text{if } Q_{k,sat} - Q_{k-1,sat} = 0 \\ Q_c & \text{otherwise} \end{cases}$

2. Computer $D_k = Q_{k,x} + Q_{k,sat} - Q_{k-1,sat}$

3. Go to the next time interval.

We applied this algorithm to the data over the two separate days (04/16/2008 and 06/26/2008) illustrated in Figure 2. The obtained matrices $D_k$ will be used as inputs for a macroscopic traffic simulation tool (FREQ12) in order to test their validity. For this purpose, we compare...
two indicators: the travel time (TT) and the total vehicle hours travelled (VHT). First, we noticed that the GCV computed by the traffic simulator were identical to the real data. This was not surprising since the same conservation law is used for both: the algorithm to retrieve the demand matrix, and the simulation tool. The most interesting results certainly pertain to TT and VHT, whose simulated figures were very close to the real data. The following reasons may explain the small gaps observed:

- the capacity of the link is given by the GFD which is a mean curve;
- the traffic simulator provides a set of fundamental diagrams among which the user needs to choose the one that is closest to his data. And yet, the shape of the provided diagrams differs from that of the diagrams usually observed on French urban motorways.

Using link-aggregated data means that the link is considered as a single section with only one entrance and one exit. Its characteristics, as required by the simulation tool, are given by the GFD, that is, its capacity, free-flow and critical speeds.

3.3. Link Capacity Function

Whether one chooses to use FREQ12, a fundamental diagram is required. However, we pointed out the imprecision of the GFD in the previous section 2.1. In the rest of this study we will continue the data analysis in order to associate the Generalised Volumes (GV and GCV) to the other link-aggregated variables, VHT and TT. For this purpose we will use the link capacity function, which relates travel times to capacity and the volume of vehicles travelling on the link (Ortúzar and Willumsen, 2009). A widely employed formulation is the Bureau of Public Roads (BPR) model, which specifies that speed decreases as vehicle flow increases:

\[
TT = TT_0 \times \left(1 + \alpha \left(\frac{V}{Q_c}\right)\right)^\beta
\]

where TT₀ is the travel time under free-flow conditions, V and Q_c are respectively the volume of vehicles present on the link and the practical capacity. Practical capacity, as defined in the first edition of the Highway Capacity Manual, is a lower volume than the possible capacity, which is chosen with the density being not so great as to cause unreasonable delay, hazard, or restrictions to drivers’ freedom to manoeuvre under prevailing roadway and traffic conditions. It usually corresponds to about 80% of the true capacity (Kutz, 2004).

From the fundamental diagram we know that low speeds may also be recorded for low flows in congested traffic. This means that the speed-flow relationship given by this diagram may serve in the calibration of the link function capacity only as long as traffic remains fluent. In this section we will propose a calibration of this function based only on the link-aggregated data.
During a given time interval, for a free-flowing steady-state traffic, flows measured by a loop detector embedded in the pavement correspond to the number of vehicles which have travelled on the link during this period. Hence, it is the volume of vehicles that are present on the link. The figure is different for congested traffic. We have already shown in the previous section that the Generalised Congested Volumes recorded on the sections of the stretch during a given period represent the storage of vehicles. This storage is the difference between traffic demand for the period (increased by vehicles already stored from the previous time interval) and the link capacity. This means that the real volume of vehicles concerned by our period is actually the sum of those that could leave the link (which is assumed to be equivalent to the link’s capacity) and those stored for the next period.

The following graphs (Figure 3) show the change in travel times with the Generalised Volumes (GV) under free-flow conditions, and Generalised Congested Volumes (GCV) under saturated conditions. The plots represent A1 data, but the same observation holds for other stretches, notably A3 and A10 motorways in the Paris region.

One may notice that the travel time remains constant when traffic is fluent, even for flows above the capacity shown by the GFD. This is likely due to the fact, already mentioned by Cassidy (2011), concerning the Macroscopic Fundamental Diagram. The capacity used here is not the actual capacity of the stretch, which would be reached if all the sections could be at the same regime simultaneously. Rather, we have a lower figure when using the GFD.

In congested traffic, we observe an increase of travel times with GCV, i.e. with the storage. Note that congested traffic over the link implies that the overall mean speed is lower than the critical one, but traffic may be fluent in some sections. The volume of vehicles that are travelling on the link corresponds to the measured traffic flow in fluid steady-state regimes, and the sum of entering traffic (demand) and the storage in congested regimes. The two graphs above may then be combined to form the link capacity function relating travel times to traffic volumes and the capacity of the stretch, as depicted in Figure 4.

Figure 3 : Travel times evolution with global volumes and volumes in congestion on A1 motorway
This section of the paper allowed us to establish a relationship between the demand matrix – obtained from the Generalised Fundamental Diagram and VKT – and traffic conditions described by travel times.

The following graphs conclude the observation of the relationships between empirical data on A1 and A3 motorways. They compare real data of VHT and TT, to simulation results with FREQ12 (see section 2.2) on the one hand, and results obtained by applying the algorithm presented in section 2.2 and the above link capacity function. We keep the same two days: 04/16/2008 and 06/26/2008.

We observed a good match between the real data and the results obtained from the joint use of the algorithm and the link capacity function. For the VKT of the two days, the total gap was +2% and -8% respectively. Differences in hourly travel times were less than 4 minutes. As for the VHT, the total gaps are 1% and 4% for the two days respectively. The macroscopic
simulation using FREQ12 recreated traffic conditions with good accuracy as well. However, the gaps with the real data were greater and we noticed a shift in the curves during congestion dissipation. As a whole, the maximum differences were observed at the beginning and end of congestion.

4. APPLICATION TO THE EVALUATION OF TRAFFIC MANAGEMENT OPERATIONS

4.1. Methodology

This part of the paper will present the application of our findings to an evaluation method of the impact of traffic management operations on congestion. The method is based on the level of service concept. Two case studies show how to take into account traffic growth in a posteriori assessments on the one hand and a capacity increase on the other.

As a rule of thumb, the impact of a traffic management operation is the difference between indicators measured at a reference situation and those measured during the post-implementation stage. Nevertheless, we need to recall two key points:
- Traffic management operations may also affect total vehicle hours travelled in uncongested regimes, but
- Only delays are monetised in socioeconomic evaluation, i.e. minutes spent in the network beyond a certain threshold usually corresponding to the critical speed.

Our proposed method is based on the levels of service, which define classes of traffic conditions arbitrarily. We use the French classification with four levels, the last one (LOS4) pertaining to congestion and for which the threshold of maximum speed is slightly below the critical speed.

Our proposed method is based on the levels of service, which define classes of traffic conditions arbitrarily. We use the French classification with four levels, the last one (LOS4) pertaining to congestion and for which the threshold of maximum speed is slightly below the critical speed.

![Figure 6: Discrimination thresholds of the levels of service on the fundamental diagram](image)

We will then assume that a vehicle travelling at the third level of service (LOS3) is not experiencing congestion even if its mean speed would be slightly below the critical speed, as illustrated in Figure 6. This consideration only concerns the French classification of the levels.
of service; this is not the case for the Highway Capacity Manual classification for instance. Since the level of service is usually employed to describe traffic locally (section-based), we will then use the expression “generalised level of service” to refer to the GFD, hence to overall traffic conditions for the whole stretch.

For each time interval in which the overall traffic flows at LOS4 in either of the two situations, the impact of a traffic management operation on traffic congestion is given by the following equation:

\[
(\delta VHT)_k = \left( TT^Y_k - TT^Y_{NSC3} \right) \cdot Q^Y_{k,x} - \left( TT^X_k - TT^X_{NSC3} \right) \cdot Q^X_{k,x}
\]

\(TT^X_{NSC3}\) and \(TT^Y_{NSC3}\) represent the maximum travel time at LOS3 for the reference (X) and post-implementation (Y) situations respectively. \(\delta VHT\) is the change in the VHT.

4.2. The Case of a Traffic Demand Growth

Traffic management usually aims at reducing the extent of congestion over a link or network. However, we may actually observe this reduction after the implementation, but along with changes in the demand level for reasons not necessarily dependant on the operation. This could happen due to an economic or demographic growth for instance. In such cases, obviously a simple before/after comparison would distort the impact assessment. In fact, even without the implementation of the operation, traffic conditions would have changed over the link or network due to the demand changes. The different elements presented in the previous sections of this paper allow us to set the reference situation for an a posteriori evaluation without using a classic simulation tool, which is usually not employed at this stage due to the amount of data required and/or available. Hence, it is possible to determine the VKT, VHT and TT that would have prevailed even if the operation had not been implemented.

The following graph illustrates a situation observed on the A1 motorway stretch presented earlier in this paper. During the spring of 2009, the inner-left lane was dedicated to taxis from 7 to 10 AM on a 4.5 km long section. After the implementation, a comparison between daily VKT for 2008 and 2009 showed a 6%-decrease in the total demand over the entire stretch. The operation did not modify the capacity of the facility, which would lead us to expect some improvement in traffic conditions. However, travel times during peak periods (7-10 AM) and (17-20 AM) slightly increased by at least 5 minutes, which represented more than 15% of the initial figures (Princeton, 2010). These observations were explained by applying our method. In fact, the demand matrices for the two situations, before (2008) and after (2009) indicate significant drops (down to -25%) in the demand level during off-peak periods, while increases ranging from +2.5% (morning) to +9% (evening) were recorded during peak periods. The results are shown in Figure 7.
The analysis of the VKT data only would not have given that insight, since this variable was systematically lower at all the time intervals in 2009. The results confirm that changes in traffic demand volumes have a greater effect on traffic conditions during peak periods, which is shown by both the fundamental diagram and the link capacity function.

This example also illustrates the weakness of the control stretch method to assess the impact of a traffic management operation on congestion. Indeed, the dedicated lane was activated during the morning period only. Let us consider the evening peak as control. Travel times have increased in both cases, but not to the same extent. For the control peak hour the increase is 30% versus 25% for the morning peak. These results would have suggested a slight positive impact of the operation on congestion, while actually the initial levels of congestion were different for the two cases, as well as the demand growth rates in 2009.

4.3. The Case of a Capacity Increase

The traffic management operation described in the previous section was implemented because of the physical restrictions of the site which did not allow any capacity increase. Now we propose to study a scenario consisting in increasing the bottleneck capacity by 5%, assuming that an auxiliary lane could be open. This 5% figure is in line with previous results obtained from this kind of operation in France, e.g. A4-A86 and A3-A86 motorway stretches (Cohen, 2005). Capacity and the travel times thresholds for the generalised levels of service are given in Table 1.

Table 1. Site characteristics (A1)

<table>
<thead>
<tr>
<th>Stretch</th>
<th>Qc</th>
<th>TTₐ</th>
<th>α</th>
<th>β</th>
<th>LOS1</th>
<th>LOS2</th>
<th>LOS3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>4300</td>
<td>9.383</td>
<td>0.35</td>
<td>5.464</td>
<td>&lt; 11</td>
<td>&lt; 12</td>
<td>&lt; 29</td>
</tr>
</tbody>
</table>

Evolution of TT and VHT

Table 2 gives travel times and total vehicle hours travelled during the morning period. For the reference situation, data are collected from the SIRIUS database. For the would-be situation with the 5% capacity increase, data are computed by applying the algorithm presented in section 2.2 and the link capacity function.
Table 2. Change in traffic conditions after 5% increase in the demand level (A1)

<table>
<thead>
<tr>
<th>A1</th>
<th>Reference situation</th>
<th>Would-be situation (+5% capacity increase)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TT</td>
<td>VHT</td>
</tr>
<tr>
<td>6H-7H</td>
<td>14</td>
<td>1242</td>
</tr>
<tr>
<td>7H-8H</td>
<td>32</td>
<td>2379</td>
</tr>
<tr>
<td>8H-9H</td>
<td>33</td>
<td>2372</td>
</tr>
<tr>
<td>9H-10H</td>
<td>29</td>
<td>1947</td>
</tr>
<tr>
<td>10H-11H</td>
<td>23</td>
<td>1732</td>
</tr>
</tbody>
</table>

These results indicate that a scheme that would increase the bottleneck capacity on A1 by only 5% would lead to about an 11-minute travel time saving during the morning peak. This would also allow cutting off congestion since the generalised level of service would change from LOS4 to LOS3.

Evaluation of the impacts on traffic congestion

For the evaluation we will only keep those time intervals during which the stretch was congested for either of the two situations (reference and would-be). This means that we will only consider time intervals between 7H and 10H AM (in blue in Table 2). As we already mentioned, only delays are taken into account in socioeconomic evaluations.

The potential impact of the operation is calculated hereafter:

- VHT in congestion at the reference situation:
  
  \[2379(1 - 29/32) + 2372(1 - 29/33) + 1947(1 - 29/29)] = 511 VHT

- VHT in congestion at the would-be situation: 0 VHT since traffic is flowing at LOS3.

The difference is then a 511 VHT daily decrease, which would represent the positive impact of the operation for the morning period. This impact may be monetised by applying the proper value of time.

5. CONCLUSION AND FUTURE STEPS

Observation of the empirical data for several motorway stretches highlight some relationships between link-aggregated data such as the total vehicle kilometres travelled (VKT), the total vehicle hours travelled (VHT) travel times (TT). This study shows that it is possible to analyse a stretch with all its access and exit ramps as a single section with one entry (the upstream end) and one exit (the downstream end). Calibration of the Generalised Fundamental Diagram (GFD) therefore allows retrieving the demand matrix corresponding to any observed congestion pattern. Application of the proposed method is straightforward in a posteriori and a priori evaluations. In the first case, it becomes possible to take into account the effect of a traffic growth on past data without using classic simulation tools that require an amount of data not always available at post-implementation stages. As for a priori evaluations, the method not only provides a rapid, and yet accurate, estimation of the impacts of a traffic management operation on traffic conditions over a stretch, but also allows...
analysing the interactions with other stretches at the network level. Furthermore, the relationship between the capacity of the stretch and the demand matrix makes it possible to analyse the impacts of several variants of traffic management operations, such as lane management, traffic routing and diversion. Since the method may be applied to all traffic regimes, it also makes it possible to assess safety and environmental impacts, within the frame of traffic management.

In this study we used the French classification of the levels of service for the evaluation of the impacts on congestion of traffic management operations. However, the highlighted relationships between the link-aggregated data may also be applied to any other arbitrary classification of the speed thresholds.

In spite of those contributions, some aspects need to be addressed more deeply. The first one concerns the detectors that equip the different sections of the stretch. It could be interesting to analyse the influence of their availability rates on the results accuracy, since we know that estimates of the link-aggregated data (VKT, VHT and TT) largely rely on the detectors performance over a considered period. Secondly, data analysed in this study pertain to typical days, i.e. working and class days, with no incident or accident and good weather conditions. It could also be interesting to test the ability of the proposed method to recreate non typical patterns, during traffic incidents for instance. This would help in determining which traffic operation may be more appropriate in an incident management.

Finally we also intend to study the influence of the temporal aggregation level of employed data on results accuracy. In fact, in this paper we used 1-hour time intervals; the next step will be to try out 6-minutes data.

REFERENCES


Cohen S., (2005).- La Gestion dynamique des voies : un outil efficace d’exploitation des
autoroutes ? (Managed lanes : an effective tool for motorways management ?) La Revue Générale des Routes, n° 842.


External Factors Affecting Motorway Capacity

Ben Morris, TRL

bmorris@trl.co.uk

1 Introduction

The concept and derivation of highway capacity originated with the seminal paper of Greenshields in 1935 (Greenshields, 1935), where the correlation between traffic speed and flow was first modelled. Subsequently, a sizeable body of research has sought to better understand the complexities of the ‘speed-flow’ relationship and to develop theoretical and mathematical methodologies for determining highway capacity that more accurately reflect the real-world dynamics of traffic flows. Whilst such research continues to contribute to effective highway design and engineering, it is importantly enabling more efficient and responsive management of traffic flows for the benefit of road users.

Traditionally, the capacity of a highway is given as a fixed value based on speed-flow relationships and geometric factors. Widely accepted design guidance, such as the Highway Capacity Manual (HCM) (Transport Research Board 2000), defines capacity as the maximum flow rate that can be reasonably expected to traverse a highway link under prevailing roadway, traffic and control conditions.

In recent years, there has been a growing recognition that capacity should not be considered as a fixed value. Empirical studies (Prevedouros & Kongsil, 2003, Chung et al, 2006) have demonstrated that capacity is dependent on external factors that vary in time and space. These external factors arise from changes in driver, vehicle, road and environmental conditions (including weather) (Ponzlet, 1996). Furthermore, even under constant external conditions, observed flow breakdowns on a particular highway can be preceded by highly variable flow rates (Brilon et al, 2005), that do not always occur at the maximum ‘capacity’ flow.

At a simplistic level, highway capacity has been calculated using a fitted distribution model so that it is equivalent to a percentage on the statistical distribution of observed flow rates. Over time, studies have refined this approach to improve accuracy of estimations, such as through the removal of long-headways data (Chang & Kim, 2000) or choosing a percentile within a specified range of observed maximum flow rates (Washburn et al, 2010). Such methods lack a theoretical basis and, without incorporation of breakdown events or congested data points, the accuracy of the estimated capacity values is unknown.

Van Arde (1995) introduced a now well-known and frequently adopted methodology for calculating capacity which is also based on the fundamental speed-flow diagram. Subsequent comparisons with other approaches have found this method to be useful in providing general capacity estimates in terms of providing a good fit with the empirical data and consistency with traffic flow theory (Washburn et al, 2010). This method incorporates congested data points which increases accuracy of estimations; however, capacity values are not tied directly to breakdown events. Therefore, it is still not possible to determine whether more traffic than the highest observed flow rates could be served.

Efforts have subsequently been made to develop a theoretical concept for incorporating breakdown events into capacity estimations, so that capacity value is based on the flows
which cause a breakdown event. The Product Limit Method (PLM) estimates a breakdown probability based on flow rates preceding breakdown events, from which capacity is then determined as a certain percentile value of the breakdown probability distribution (Brilon et al, 2005). This method was only applied to a highway with a bottleneck where the lane “drops” (Brilon et al, 2005), but later studies validated that PLM can be applied more widely to basic highway links for estimation of capacity values (Washburn et al, 2010). It should be noted that the determination of appropriate breakdown probabilities is complex, and further studies (Washburn et al, 2010) suggested that estimates were only reliable for sites with more than 0.5 breakdown events in one day.

In the studies referenced above, it is commonly noted that estimated capacity rates are lower than those predicted by the HCM. It is thought that this is explained in part by the increasing use of more disaggregated data values, as well as the understanding that external factors can contribute to flow breakdown well below the ‘ideal capacity’. Additional guidelines (NCHRP, 2008) provide equations to enable improved application of HCM methods in planning applications that take local factors into account.

It is only relatively recently that more research has been undertaken on the impact of weather on highway capacity. Prevedouros & Kongsil (2003) provide a synthesis of 26 previous studies investigating the effects of wet conditions on highway speed and capacity. The results from 11 studies (post-1980), where original data from highways are averaged assuming equal weights, produced an average speed reduction of 4.7mph in light rain. This is significantly less than the 12.0mph decrease suggested in the HCM (Transportation Research Board 2000). The average capacity reduction is 8.4% in light rain (seven studies).

A more recent study (Chung et al, 2006) was undertaken in Japan, and used high resolution weather and traffic data to investigate the impact of rain on motorway capacity and speed, and the effect of daylight on capacity. Findings demonstrated that rain decreases capacity ranging from 4 – 7 % for light rain and 14% for heavy rain. This is accompanied by a noticeable decrease in speed (between 4.5% and 8.2% in light and heavy rain conditions respectively). Capacity during daylight was found to be 12.8% greater than capacity during daybreak. Due to the limited sample size, as only one site was used, Cung et al were not able to draw firm conclusions. However, the results regarding capacity reduction roughly corresponded with previous studies (Prevedouros & Kongsil, 2003).

It is clear that due to a lack of studies, or in some cases due to a lack of consistency in findings between studies, there is still extremely limited understanding of the effect and level of impact of external factors on highway capacity. The majority of methodologies rely on a calculation of capacity as a percentile of observed flow rates although, as mentioned, the PLM method using a stochastic estimation model does allow investigation of external factors whilst also considering the ‘random’ element of capacity variance. Research (Brilon et al, 2005) has also highlighted that cultural differences in driver behaviour can play an important part in accounting for variance, making it difficult to compare international findings and highlighting the importance of conducting UK studies to inform applications for the UK network.

The purpose of this paper is to investigate factors that affect capacity, and in doing so, build a simple statistical model that parameterises these factors and describes the variable nature of capacity. This paper goes further than existing studies through
incorporating multiple capacity affecting factors to build a fuller picture of what causes capacity to vary.

2 Methodology

2.1 Choosing factors to investigate

Many factors are thought to affect motorway capacity. Table 1 below includes a non-exhaustive list of potential factors that affect motorway capacity that was developed as part of this research.

<table>
<thead>
<tr>
<th>Road Geometry</th>
<th>Driver Behaviour</th>
<th>External</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lane width</td>
<td>Lane changing</td>
<td>Rainfall</td>
</tr>
<tr>
<td>Link length</td>
<td>Gap acceptance</td>
<td>Visibility</td>
</tr>
<tr>
<td>Gradient</td>
<td>Headway adjustment</td>
<td>Time of day</td>
</tr>
<tr>
<td>Curvature</td>
<td>Speed variation</td>
<td>Day type</td>
</tr>
<tr>
<td></td>
<td>Use of brakes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Traffic composition</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Traffic demand</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Commuter, Commercial or Leisure Traffic</td>
<td></td>
</tr>
</tbody>
</table>

The parameterisation and integration of all these factors into a unified model was beyond the scope of this project. A subset of these factors was therefore shortlisted for investigation. The criteria and process for selecting which factors to assess is outlined below:

- Factors should be easy and simple to parameterise and measure.
- Factors should have readily available data sources to allow analysis to be performed on the associated parameters.
- Geometric factors were not considered for this project as they have been covered extensively in the HCM (Transportation Research Board 2000).

The shortlist based on the criteria above is included in Table 2 below.

<table>
<thead>
<tr>
<th>Internal</th>
<th>External</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merge Percentage</td>
<td>Rainfall</td>
</tr>
<tr>
<td>Diverge Percentage</td>
<td>Visibility</td>
</tr>
<tr>
<td>Heavy Goods Vehicle Percentage</td>
<td>Light, Twilight, or Night</td>
</tr>
</tbody>
</table>
On the UK Strategic Road Network, the Highways Agency collect extensive data from inductive loops embedded in the road surface. This data, collected from the Motorway Incident Detection and Automatic Signalling network, is referred to as MIDAS data for short. MIDAS data does not include lane changing information, as the loops are static in space and are spatially too infrequent to infer lane changing from in-lane data. It was therefore decided that the parameters of merge percentage and diverge percentage be considered as a proxy for lane changing. This is based on the evidence that most flow-breakdown occurs in close proximity to merge and diverge junctions, and very rarely mid-link.

MIDAS data also includes length classification information. This meant that traffic composition could be easily parameterised. Therefore Heavy Goods Vehicle flow as a percentage of main carriageway flow was also considered so that an assessment could be made of how the proportion of HGVs might affect capacity.

Wunderground, an online weather information provider, catalogues historical weather data from its weather station network. Precipitation information, recorded by various weather stations was referred to during the analysis that was performed in this project.

2.2 Choosing sites to investigate

The criteria for selecting sites, therefore, were as follows:

- Congestion must be recurrent, demonstrating flow-breakdown on at least 3 days a week on average.
- MIDAS data must be available for at least a year, both at the immediate location of the congestion seed point, as well as upstream and downstream of the surrounding junctions.
- MIDAS data must also be available on the surrounding slip roads.
- Weather information must be available from a representative site within a 5 mile radius of the congestion seed point.

Based on these criteria, congestion seed points on the UK Strategic Road Network (SRN) were catalogued and categorised. From this list, the most suitable sites were then selected. These sites were:

1. The M4 from junctions 19 to 20 in the eastbound direction.
2. The M25 from junction 11 to 12 in the anti-clockwise direction.

2.3 Data description

The raw data used in this study consists of:

- MIDAS Data (from the Highways Agency)
- Historical weather data (from Wunderground)

MIDAS data contains the following:

- 1-minute averaged speeds by lane
1-minute averaged occupancy by lane
1-minute traffic counts by lane (normalised to hourly flows)
1-minute classified vehicle counts by carriageway

Wunderground weather data simply recorded the normalised hourly equivalent rainfall, where the measurement period varied from station to station and also within station.

Once two appropriate sites were selected, data for each site was then collected. To build data sets that would then be used for statistical analysis, MIDAS data for 2008 and 2009 was collected for each site. From this data two operations were performed:

1. Visual identification of congestion seed points and shockwaves using space-time plots generated from TRL’s Motorway Traffic Viewer (MTV) software.
2. Data processing to extract the relevant data for the 15 minute period prior to the onset of flow breakdown.

### 2.4 Data collection and handling

MTV is a tool that has been widely adopted in the UK and the concepts it uses in terms of traffic data visualisation have been even more widely adopted in the academic community. An example of an MTV space-time plot is shown in Figure 1.

![MTV plot](image)

**Figure 1 – An MTV plot showing the standing wave (seed point) and resulting shockwaves**
Using this tool, each day for 2008 and 2009 was examined. The onset of flow breakdown can be identified as the left most part of the standing wave of low speed vehicles. The seed point is also identifiable through traffic immediately downstream travelling away from the standing wave at free-flow speeds. Through this method of seed point identification, the time at the start of flow breakdown was recorded to the nearest 5 minutes.

Once this initial data set was captured, a fully automated set of analysis was then performed using Microsoft SQL Server, which hosted all the MIDAS data that had been collected. Table 3 outlines how data was handled for each parameter.

### Table 3 – Description of how the data was handled

<table>
<thead>
<tr>
<th>Operation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finding the peak 15-minute throughput</td>
<td>The times at which flow breakdown occurred were derived from visual inspection. Inherent in this process is a lack of fineness. To improve the granularity, data from plus and minus 15 minutes of the initial times identified was analysed, where the maximum 15-minute throughput was extracted on a 1-minute rolling sum.</td>
</tr>
<tr>
<td>Calculating merge and diverge flow</td>
<td>As described in the site selection criteria above, sites had MIDAS data available on the slip roads as well as upstream and downstream of the adjacent merge and diverge slip roads. These data were used to derive the merge and diverge flows based on the difference between the flows upstream and downstream of the merge and diverge slips. These flows were then validated against the slip road flows for completeness. These flows were then described as a percentage of the 15-minute throughput prior to flow breakdown.</td>
</tr>
<tr>
<td>Calculating HGV flows</td>
<td>MIDAS data stores classified vehicle counts. The longest category (&gt;11.6m) was used and vehicles in the class were considered as HGVs for the purpose of this study. These counts were then described as a percentage of the 15-minute throughput prior to flow breakdown.</td>
</tr>
<tr>
<td>Precipitation</td>
<td>Normalised hourly precipitation values were looked up for the 15-minute throughput prior to flow breakdown.</td>
</tr>
<tr>
<td>Night, Twilight or Daylight</td>
<td>These values were looked up for each day that flow breakdown had occurred. The relevant state was then recorded based on the category in which the greatest period of the 15-minutes prior to flow breakdown fell.</td>
</tr>
</tbody>
</table>
3 Results

The data sets generated as described in the section above were subjected to statistical analysis of variance (ANOVA) tests using the following factors:

- Merge flow as a percentage of main carriageway flow
- Diverge flow as a percentage of main carriageway flow
- HGV flow as a percentage of main carriageway flow
- Precipitation as a categorical variable (wet or dry) and was limited to the 15-minute period prior to flow breakdown.
- Lighting conditions as a categorical variable (darkness, morning twilight, daylight, evening twilight)

The analysis of variance tests each factor for statistically significant contributions to explaining the observed variance in capacity. Partial eta squared is a value between 0 and 1 that describes the relative contribution of each parameter to explaining the variance. F describes the ratio of explained variances between terms and the significance term indicates the probability that the correlation is due to chance.

The results for the M4 J19 – J20 ANOVA are detailed in Table 4 below.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>F</th>
<th>Significance</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>9</td>
<td>21.188</td>
<td>0.000</td>
<td>0.463</td>
</tr>
<tr>
<td>Intercept</td>
<td>1</td>
<td>255.263</td>
<td>0.000</td>
<td>0.536</td>
</tr>
<tr>
<td>HGV percentage</td>
<td>1</td>
<td>74.502</td>
<td>0.000</td>
<td>0.252</td>
</tr>
<tr>
<td>Diverge percentage</td>
<td>1</td>
<td>6.728</td>
<td>0.010</td>
<td>0.030</td>
</tr>
<tr>
<td>Merge percentage</td>
<td>1</td>
<td>36.122</td>
<td>0.000</td>
<td>0.140</td>
</tr>
<tr>
<td>Precipitation</td>
<td>1</td>
<td>9.816</td>
<td>0.002</td>
<td>0.043</td>
</tr>
<tr>
<td>Lighting</td>
<td>3</td>
<td>5.794</td>
<td>0.001</td>
<td>0.073</td>
</tr>
<tr>
<td>Precipitation * Lighting</td>
<td>2</td>
<td>0.127</td>
<td>0.881</td>
<td>0.001</td>
</tr>
</tbody>
</table>

For categorical variables, lighting and precipitation in this case, the ANOVA tests for interaction between them, where if there is a high degree of interaction, the terms are likely to be dependant or co-variant with a confounding factor.

In this case, the model explains 46.3% (partial eta squared is 0.463 and is measured between 0 and 1) of the variance in observed capacity.

All factors investigated at this site are statistically significant. The HGV percentage contributes the greatest impact to the variability of capacity, followed by merge percentage. Diverge percentage, precipitation and lighting, despite their statistically significant contribution, all have relatively little impact on the variability in capacity with respect to HGV percentage and merge percentage. The ANOVA results for the M25 J11 – J12 are also included in Table 5 below.
At this site, the model accounts for 27.4% of the explained variance (partial eta squared is 0.274 for the whole model).

In comparison to the M4 site, not all factors are statistically significant in contributing to the explained variability in capacity. Both merge percentage and precipitation have no statistically significant contribution to the explained variance in capacity. Interestingly, the greatest contributor to the variability in capacity is the diverge percentage. HGV percentage and lighting contribute relatively little compared to diverge percentage, but their contribution to explaining the variability in capacity remains statistically significant.

Based on the analysis that has been performed at both study sites, the factors that affect capacity appear to be site specific. This is in itself an interesting finding that could be further verified by extending this analysis to more sites. It is also encouraging that statistically significant correlations can be drawn from relatively aggregated data, which would imply that more granular and higher quality data could further improve the explained variance in capacity.
4 Discussion

The original hypothesis at the outset of this study was that a linear equation could be constructed that could describe the variability in capacity that is observed at many points on the UK SRN, using readily available traffic and weather data. While that aim has not been completely fulfilled the results have demonstrated that there is some merit in the novel approach that has been developed in this study.

The results have also shown that there is an interaction of factors that is site specific in a way that is not yet clearly understood, but that can potentially be measured and that number of factors and/or the resolution of data are not sufficient to fully describe the variance in capacity but that they can describe a proportion of the variance with statistical confidence.

Analysis of the results shows that at each location and for each type of analysis, only the merge or diverge flows are significant, with the respective parameter showing a lesser degree of significance. This would indicate that flow breakdown is contributed to from lane changing due to either merging traffic, or diverging traffic, not both.

There is also a noticeable difference in the amount of capacity variance explained by the analysis at each site, where less variance is explained for the M25 site than the M4 site. This indicates that the parameters investigated do not account for any site specific factors. It is not known whether these can be purely explained through geometric differences or if there are other factors that have not been taken account of.

The relative significance of HGVs is also different for both sites and even contributes to an increase in capacity on the M25. This is a counter intuitive effect and could form the basis of further study.

The novelty in the approach that was taken for this study is twofold. Firstly, capacity has long been considered constant and that constant is defined somewhat loosely as the flow which can be sustained under prevailing conditions. Secondly, the approach was developed to test if readily available data could be used to test the original hypothesis above, thereby avoiding costly instrumentation and data collection regimes. This study has re-cast capacity as a variable property, which is determined by the interaction of factors prior to the onset of congestion.

The values of partial eta squared obtained for each site demonstrate that, to a varying degree, a significant proportion of the variability in capacity can be explained using the simple methods developed in this study. However, there are significant shortcomings in the methods and data that were employed in this study, due to the constraints that were imposed at the outset, namely, the requirement that only readily available data was to be used.

Relationships that cannot be verified or tested using this methodology are listed below. This list is followed by a further list describing potential improvements and next steps that could be taken to improve on the results obtained in this study.

Relationships and concepts not testable or verifiable by using this technique:

- The use of merge and diverge flows as proxies for lane changing
- How factors interact to vary the observed capacity
- The physical mechanisms that induce congestion at a microscopic level
Discrimination between types of driver (commuter or leisure traffic for instance) and any effect they might have on the variability of capacity

How geometric factors affect capacity

Further enhancements that could improve and refine the method and results generated in this study could be:

- Perform analysis on more sites
- Collect higher resolution data
- Collect better quality data
- Introduce more parameters
- Improve the rigor in defining congestion seed points

This study has sought to investigate factors that affect motorway capacity. Factors were tested to determine the statistical extent of their contribution to explaining the variability in capacity. Merge, diverge and HGV percentages, as well as lighting conditions and precipitation were all found to contribute to the variability in capacity although each factor affected capacity differently depending on the motorway location. This research has been considered a success and could form the basis of further, more detailed study that could include more granular, higher quality data as well as more motorway sites.
References & Bibliography


RELIABILITY OF IN-VEHICLE WARNING SYSTEM FOR RAILWAY LEVEL CROSSINGS – A USER-ORIENTED ANALYSIS
ABSTRACT

The paper will analyse the reliability of an in-vehicle warning system developed in Finland during years 2008-2010. The system is based on the positioning of trains using GPS, calculation of the states of level crossings on a server and in-vehicle equipment which retrieves information about the states of level crossings from the server. The in-vehicle system also detects the situations in which the equipped car approaches a level crossing. Information about the reliability of the system is very relevant for accurate estimation of the impacts of the system and for estimating potential for improvements to the system. The paper starts with a description of the system under analysis, continues by defining the concept of reliability and provides an estimate for the reliability of the system from user point of view. To achieve this objective, the paper will define the relevant concepts and describe a methodology for analysis of the reliability of the system. The main inputs to the analysis of reliability will be a short overview of existing concepts related to reliability of in-vehicle ITS systems and empirical data obtained in a field test currently running in Southern Finland.

1. INTRODUCTION

Background
The safety of level crossings is a major safety issue for railways and an existing traffic safety problem for road users. First, the active and passive safety features of most cars have not been designed to protect the vehicle occupants in a collision with a heavy and possibly fast moving train. Second, several countries still have unprotected level crossings on their railway networks. For example, the Finnish railway network has over 4000 level crossings in total, and over 3000 of them are unprotected level crossings with no bells, lights or gates. On average, close to 10 people die in level crossing accidents every year in Finland and about 50 are injured. At least some of these accidents could be avoided, if the alertness of the road user could be increased. For example in 2009, there were eight fatal level crossing accidents in Finland, and all except one of them occurred at unprotected level crossings. Only one of the eight fatal accidents in 2009 occurred at a level crossing with gates but that accident was found to be a suicide.

The number of people injured or killed in road accidents depends on three factors: exposure, accident rate and injury severity (Elvik and Vaa 2004). The in-vehicle warning system for level crossings aims to reduce the accident rate by increasing the alertness and situation-awareness of the road user in situations in which the road user is approaching a railway level crossing. Among other things, the impact of the system on the behaviour of road users depends on the reliability of the system.
In-vehicle warning systems for railway level crossings have been under development in recent years at least in Finland, the USA and Australia. The functionality and operating principle of the system analysed in this paper has been described on a general level elsewhere (Öörni and Virtanen 2007) after a small-scale proof-of-concept test carried out in 2006. The system is based on the positioning of trains using GPS, calculation of the states of level crossings on a server and in-vehicle equipment which retrieves information about the states of level crossings from the server and detects the situations in which the equipped car approaches a level crossing.

Objectives of the paper
The objective of the paper is to analyse the reliability of an in-vehicle warning system for railway level crossings developed and tested in Finland during 2008-2011 from the user point of view. This objective will be achieved by collecting data with a field test including observations on the output of the system, analysing the results and providing a quantitative estimate for reliability of the system. The reliability of the system was analysed from the user point of view for two main reasons. First, analysing the reliability experienced by end-users gives an overall picture of the current reliability level of the system and provides guidance on the factors contributing to it. Second, the reliability experienced by end-users largely determines the potential the system has to improve the safety of road users. Third, having a definition for reliability from earlier research also helped to define the scope of the paper.

Overall approach followed
The work was started by first defining the concepts and tools needed to analyse the reliability of the system and then providing an overview of the technical architecture of the system as a simple chart. The empirical data needed to analyse the reliability of the system was collected with video monitoring and with the data logging features implemented in the system. The final part of the work was the analysis of the reliability of the system with the methods described in the paper using the data collected in a field test.

2. SYSTEM DESCRIPTION

Functionality of the implemented system
The basic functionality of the system is to provide a warning of an approaching train to a car driver when he or she is approaching a level crossing. The in-vehicle system has the coordinates of all level crossings as a table and a GPS receiver. When the vehicle moves on the road network, the in-vehicle system continuously calculates distances between the vehicle and the nearest level crossings using the results of GPS positioning and the coordinates of the level crossings to detect the situations in which the car approaches a level crossing. After an
approaching situation has been detected, the in-vehicle device sends a query to the server with
the number of the level crossing the vehicle is approaching. The server continuously receives
real-time information on the positions of trains on the rail network and calculates the status of
level crossings on the basis of that information. When the server receives a query with the
number of a particular level crossing, it responds with the current status of the level crossing.
Any level crossing may have four different states: “no information”, “coming”, “alarm” or
“passed”. When the in-vehicle system has detected a situation in which the vehicle is
approaching a level crossing and the status of the level crossing in the response received from
the server is “alarm”, the in-vehicle system warns the driver with visual indication and by
playing a sound. An overview of the architecture of the system is presented in Figure 1.

![Figure 1. In-vehicle warning system for railway level crossings – simplified technical
architecture (Öörni and Virtanen 2007).](image)

The warning functionality provided by the system is based on both distance and the direction
in which the vehicle is travelling. If the distance between the vehicle and level crossing is less
than 1500 metres but more than 200 metres and the vehicle is travelling towards the level
crossing, the in-vehicle device warns the driver with a short sound and a visual indication. If
the distance between the level crossing and vehicle is less than 200 metres, the direction of
travel has no effect and the system warns with a continuous audible sound and visual
indication if the status of the level crossing is “alarm”.

**Functional requirements**
The high-level operating requirements for the system were divided into two categories: basic
functionality and other requirements. Three main requirements related to the basic
functionality of the system were identified:
- The driver receiving a warning must have enough time for braking
- Provision of a clear warning the meaning of which is easy to understand
- Provision of relevant information: the right information in the right place at the right
time to the right users

Two main requirements in the category “other requirements” were identified:
- Ease of use
- Possibility to implement the in-vehicle part of the system on the same platform with other ITS applications

The functional requirements for the system in terms of driver reaction time were defined to be similar to the functional requirements set for level crossing warning systems in Finland.

**Technical architecture**

The technical architecture of the system used in the field test is illustrated below (Figure 2) to provide an overview of the way the system has been realised and the potential factors contributing to its reliability. The level crossing server calculates the status of the level crossings on the basis of real-time positioning information received from trains equipped with MC40, CL341 and A1 Trax devices.

![Technical architecture of the in-vehicle warning system for railway level crossings used in the field test.](image)

The in-vehicle system used in the field test was a commercially available navigator phone Nokia 6710 Navigator with Symbian 9.3 operating system and support for Java midlets. The phone has a built-in GPS receiver and capabilities for GPRS and UMTS data transmission. It also uses the Nokia AGPS assisted GPS service to improve positioning accuracy and shorten the time between a cold start and first positioning fix. The level crossing server was a desktop computer running Windows 2000, a level crossing server application, the MySQL database engine and a PHP interpreter.
Three types of positioning devices installed in trains were used in the field test: A1 Trax, CL341 and MC40. It is a tracking and security unit designed for in-vehicle use in road transport. A1 Trax is equipped with built-in GPS and is capable of GPRS communication. CL341 is an in-vehicle telematics unit that has a built-in GPS receiver but no GPRS modem. Therefore, an external GPRS modem (MC35i) had to be used with CL341. MC40 is an in-vehicle telematics unit designed for road traffic applications. MC40 has a Windows CE operating system and an internal GPRS modem and GPS receiver. A1 Trax devices communicated with the level crossing server using TCP as a transport protocol for positioning messages. MC40 and CL341 communicated with the mobile gateway using a proprietary transport protocol different from TCP. The positioning data transmitted by CL341 and MC40 devices was received by a dedicated mobile gateway server and then forwarded to level crossing server using a web service interface.

3. METHODS FOR ANALYSING RELIABILITY

Reliability as a concept
Reliability is a concept that has different definitions in different contexts. Thus a clear definition of reliability is presented here. The evaluation framework for in-vehicle safety systems defined in a study on the design and validation of advanced driver assistance systems (Gietelink 2007) defines reliability in a way originally mentioned (Storey 1996) as “the probability of a component, subsystem of complete system to function correctly over a given period of time under a given set of operating conditions”. The study by Gietelink (2007) also mentions indicators for reliability applicable to warning systems and their definitions and focuses on the system reliability from the human point of view: the performance of the system visible to the user in terms of true positive, true negative and false positive and missed alarms (Table 1).

Table 1. Prediction matrix with the number of samples, categorizes as true negatives, true positives, false negatives or false positives (Gietelink 2007), adapted originally from Lee and Peng 2005.

<table>
<thead>
<tr>
<th>Actual data</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negative (safe)</td>
</tr>
<tr>
<td>Prediction</td>
<td></td>
</tr>
<tr>
<td>Negative (safe)</td>
<td>$N_{TN}$</td>
</tr>
<tr>
<td>Positive (threat)</td>
<td>$N_{FP}$</td>
</tr>
</tbody>
</table>

These four variables can be used to calculate various reliability measures (table 2).

Table 2. Reliability measures calculated from the prediction matrix (Gietelink 2007).
The use of probabilistic techniques such as the methods described in the tables above is problematic with events that occur relatively rarely such as potentially hazardous situations that occur only in certain conditions, or failures of the system for which the mean time between failures (MTBF) is longer than or equal to the observation period. However, even if the results are not statistically significant, they can provide useful information about the system.

**Reliability block diagrams**
Reliability block diagrams (U.S. Department of Defence 2004) have long been used to analyse the reliability of large and complex systems. A reliability block diagram presents a system as a group of components which may be functionally parallel to or in series with each other. Reliability block diagrams are most useful when the different parts of the system are independent of each other. In other words, this refers to a situation in which the failures of the different parts of the system are not correlated with each other. Reliability block diagrams can be used together with probabilistic techniques or as tools to understand the behaviour and topology of the system under analysis.

**QUANTIS – Quality assessment and assurance methodology for traffic data and information services**
QUANTIS – Quality assessment and assurance methodology for traffic data and information services (QUANTIS 2011) was a European project which developed a comprehensive framework for the analysis of data and service quality in ITS applications. The framework has five separate elements for data quality: availability, integrity, veracity, precision and timeliness. The elements of data quality are represented by individual quality attributes that may have quantitative measurable values. The quality elements and quality attributes present in the QUANTIS approach have been documented in the first deliverable of QUANTIS (Öörni et al. 2009). However, the framework developed in QUANTIS has no single definition.
for reliability and was originally developed to analyse and improve the quality of traffic information services.

**Method chosen**
The reliability of the system was defined as the reliability visible to the end-user according to the definition used by Gietelink (2004). A reliability block diagram was drawn to illustrate the service chain and physical and logical components contributing to the reliability of the system. Because no directly applicable definition for a successful use case or acceptable service quality was available in earlier research, concepts available in QUANTIS and information related to level crossing warning equipment were used to establish criteria for a successful use case and to classify the other possible outcomes of a situation in which a train passes a level crossing monitored in the field test.

### 4. MODELLING OF THE SYSTEM

**Modelling reliability**
A reliability block diagram (Figure 3) of the system was drawn on the basis of the technical architecture (Figure 2). The diagram illustrates the components affecting the reliability of the system.

![Figure 3. Reliability block diagram of an in-vehicle warning system for railway level crossings.](image)

The system implemented in the field test has three types of train equipment: A1 Trax, CL341 and MC40. The latter two send their positioning messages to the level crossing server via mobile gateway, instead of sending them directly from the train to the level crossing server as A1 Trax devices do. For this reason, two separate scenarios have to be distinguished in the
reliability block diagram: a rail bus or diesel engine equipped with a A1 Trax device approaching the level crossing or a diesel engine equipped with a CL341 or MC40 device.

Data collection for the reliability analysis was carried out at three points in the service chain: (A) where train position information is received by the level crossing server directly from either the train equipment or mobile gateway, (B) at the interface providing level crossing status information to in-vehicle device, and (C) between the in-vehicle system and end user. Data collection at points A and B was carried out automatically using the data logging features built into the level crossing server software. Data collection at point C was carried out by monitoring the level crossing and the in-vehicle system with video cameras and using a multiplexer to combine the two video streams into the same video data file. Because the focus was on studying reliability from the user point of view, the reliability analysis carried out in this paper was performed at point C.

The interface between the in-vehicle system and GPRS and public internet was not monitored because no easy-to-use data logging tools were available for the mobile hardware and software platform used in the field test, and it was assumed that the added value of such information would be limited because of the extensive data collection in other parts of the service chain.

The definition for a successful alarm was created on the basis of the user requirements and the basic functionality the system should provide. In other words, the alarm provided by the in-vehicle equipment to the driver should start early enough and continue without interruption until the train has arrived at the level crossing. The correct alarm belonging to category $N_{TP}$ and various types of unsuccessful alarms are illustrated below (Figure 4). The other alarms in the picture are classified in category $N_{FN}$. Even though the system has in fact provided a warning in those cases, classifying these situations as “false negatives” can be reasoned as follows: a warning coming too late, ending too early or being interrupted is obviously outside the stated quality boundaries and may be useless for the driver or cause confusion.
Criteria for a successful alarm were based on the main requirements set for the system and those set for fixed warning equipment at level crossings. The latest acceptable time for starting the alarm must give the driver enough time to react and stop the vehicle; 25 seconds was assumed to be enough \((y = 25 \text{ s})\). The earliest acceptable time for starting the alarm must ensure that the information the system provides is useful to the driver. Four minutes was considered to be the maximum time from the alarm to arrival of the train at the level crossing \((x = 4 \text{ min})\). The latest acceptable time for ending the alarm must ensure that warnings that are no longer relevant do not confuse users and reduce their trust in the system. The latest acceptable time for ending the alarm was assumed to be 20 seconds after the train has passed the level crossing \((z = 20\text{ s})\).

**Experimental setup**

Information about warnings provided by the system and actual trains passing the level crossing was collected with video monitoring. Video streams from two cameras aimed at the display of the in-vehicle system and the level crossing were combined into the same picture and recorded on a hard disk together with a timestamp generated by the hard disk recorder. The collected video data was coded into mpg format with 25 Hz frame rate and viewed with Avidemux open source video processing software. Various types of events such as all trains that passed the level crossing and all alarms given by the in-vehicle system were documented as a spreadsheet. In total, 500 gigabytes of video data were collected between 7th and 15th March. The data collection was interrupted only when the battery of the monitoring system was changed every day to a fully charged one and when the monitoring system was moved from the Lappohjan satama to Skogbyn seisake level crossings.

The system used for data collection is presented in Figures 5 and 6. Video monitoring was carried out at two level crossings: Lappohjan satama (from 7th to 15th March) and Skogbyn seisake (from 15th to 18th March). The monitoring system with the in-vehicle device was...
installed on an electric pole about 50-100 metres from the level crossing at both the Lappohjan satama and Skogbyn seisake level crossings.

Figure 5. Overview of the system used for video data collection.

Figure 6. Video monitoring equipment used for video data collection.

Lappohjan satama (Figure 7) is a level crossing between Santala and Skogby train stops on a railway line between Hanko and Karjaa in Southern Finland. The level crossing has two railway tracks and three types of railway traffic: passenger trains operated as a rail bus between Hanko and Karjaa, freight trains from the northeast to Hanko harbour or to Koverhar steelworks or vice versa, and shunting activities related to a branch line leading to the nearby Koverhar steelworks. The railway line between Hanko and Karjaa is a non-electrified and mostly a single-track line.
Skogbyn seisake level crossing (Figure 8) is located just adjacent to the Skogby train stop in the Raasepori municipality. The railway is a single-track line at that spot. Rail buses but not freight trains stop at the Skogbyn seisake train stop. There are no shunting activities close to that level crossing because the location is several kilometres from shunting yards and railway stations handling freight.

5. RESULTS AND DISCUSSION

Reliability results
The video monitoring results are summarised in Table 3.
Table 3. Results of video monitoring in the field test site.

<table>
<thead>
<tr>
<th>Passed trains</th>
<th>No quality deviations</th>
<th>Quality deviations</th>
<th>No train</th>
<th>Missed alarm</th>
<th>False alarm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Late start</td>
<td>Interrupted</td>
<td>Late start + interruption(s)</td>
<td>Late end</td>
<td>Early start</td>
</tr>
<tr>
<td>7.3.2011</td>
<td></td>
<td></td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.3.2011</td>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00-07:46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.3.2011</td>
<td></td>
<td></td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>07:46-13:38</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.3.2011</td>
<td></td>
<td></td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13:38-24:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.3.2011</td>
<td></td>
<td></td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.3.2011</td>
<td></td>
<td></td>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.3.2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00-12:17</td>
<td></td>
<td></td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.3.2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:17-24:00</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.3.2011</td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.3.2011</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.3.2011</td>
<td></td>
<td></td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>15.3.2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>16.3.2011</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.3.2011</td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.3.2011</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results of video monitoring have been added to the prediction matrix below (Table 4). A large number of missed alarms occurred early in the test when the service was not operational because of a server failure. Failures of the level crossing server are relatively rare events that do not usually occur within a period of 11 days. For this reason, the measured data does not accurately reflect the impact of that failure mode on the reliability of the system. The time during which the server was not operating (6th March at 22:36 – 8th March at 7:46 and 8th March at 13:38 – 11th March at 12:17) was excluded from the observation period and any observations recorded during that period were not included when the figures in Table 4 were calculated.
Table 4. Prediction matrix with number of observations categorized as true negatives, true positives, false negatives or false positives derived from observations in Table 3.

<table>
<thead>
<tr>
<th>Actual data</th>
</tr>
</thead>
<tbody>
<tr>
<td>No train</td>
</tr>
<tr>
<td>Prediction</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Table 5. Reliability measures calculated from the figures in Table 4.

<table>
<thead>
<tr>
<th>Rate</th>
<th>Definition</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real occurrence rate $p$</td>
<td>$(N_{FN} + N_{TP})/((N_{TN} + N_{FP} + N_{FN} + N_{TP})$</td>
<td>92.7%</td>
</tr>
<tr>
<td>Accuracy $p_{\text{accuracy}}$</td>
<td>$(N_{TN} + N_{TP})/((N_{TN} + N_{FP} + N_{FN} + N_{TP})$</td>
<td>35.2%</td>
</tr>
<tr>
<td>Precision $p_{CP}$</td>
<td>$N_{TP}/(N_{FP} + N_{TP})$</td>
<td>82.9%</td>
</tr>
<tr>
<td>True positive rate $p_{TP}$</td>
<td>$N_{TP}/(N_{FN} + N_{TP})$</td>
<td>38.0%</td>
</tr>
<tr>
<td>False negative rate $p_{FN}$</td>
<td>$N_{FN}/(N_{FN} + N_{TP})$</td>
<td>62.0%</td>
</tr>
<tr>
<td>True negative rate $p_{TN}$</td>
<td>$N_{TN}/(N_{FN} + N_{FP})$</td>
<td>0% (not applicable)</td>
</tr>
<tr>
<td>False positive rate $p_{FP}$</td>
<td>$N_{FP}/(N_{FN} + N_{FP})$</td>
<td>1 (not applicable)</td>
</tr>
<tr>
<td>Reliability $p_{rel}$</td>
<td>$\sqrt{(N_{TP}^2 / (N_{FP} + N_{TP})(N_{FN} + N_{TP})}$</td>
<td>56.1%</td>
</tr>
</tbody>
</table>

The share of false positives of all alarms generated by the system is 17.1%.

**Limitations of the method**

Video monitoring is a labour-intensive data analysis method. For this reason, only a relatively short observation period of 11 days was feasible. Thus, the results have to be interpreted with caution for all events for which the observation period is shorter or of the same length as the mean time between a particular type of failure (MTBF) or the time to repair it (MTTR). For this reason, the impact of server failures which occurred early in the observation period also had to be excluded from the data to give more accurate information on the reliability of the system in typical operating conditions.

The system analysed in this paper is a large and complex one that relies on other systems such as mobile and fixed-line communication networks and satellite positioning. This means that there may be failure modes that did not appear within a short observation period of 11 days and in two physical locations. However, the results most probably reflect the impact of failure
modes which occur most frequently and can be expected to have the greatest impact on the reliability of the system.

6. DISCUSSION AND CONCLUSIONS

Concluding remarks
The true positive rate was 38.0% if only alarms meeting the previously defined criteria (63 alarms) are classified as successful alarms (Table 4). If also alarms with quality deviations (an additional 41 alarms) are counted as true positives, the true positive rate improves to 62.7%. In terms of the reliability metrics calculated in Table 4, the system tested in the pilot does not meet the reliability level acceptable for a commercially available ITS service in its present form.

Many observed situations classified as “false negatives” are actually realised detections with some deviations from the quality criteria set earlier in the paper and hence not “totally” missed, and some of them probably still provide useful information to the driver. While some of them may be hazardous to the user in certain situations, others are only a nuisance such as alarms not ending quickly after the train has passed. It is also possible that the user may see a certain amount of failures as acceptable for a new application. Some types of quality deviations may also be more and others less acceptable from the user point of view. In any case, the impact of various types of quality deviations on the end users should be assessed to determine which of them are most likely to have serious consequences.

The system generated only 13 false alarms during the observation period, which is a relatively small amount when compared to 63 successful alarms and 166 trains observed in total (Table 4). Finding a balance between a low false alarm rate and a high probability of detection is a common objective but also a challenge in the design of alarm systems.

The test shows that the expected functionality has been realised by the system, but the reliability level of the pilot system must be improved to meet the requirements of a commercially available ITS service. The results indicate a moderate false alarm rate and a large false negative rate. The challenge in the further development of the system seems to be the share of false negatives rather than false positive detections.

Once the results were available, preliminary analysis of the causes of unsuccessful alarms was performed, focusing on situations in which the system had provided no detection at all (56 events, Table 3). Twelve of the missed alarms were found to be have been caused by an engine, rail bus or work machine passing the level crossing without a train unit connected to the system. An additional 15 missed alarms were caused by handovers from one GSM base
station or network cell to another, or by other situations in which the data connection between the train unit and the level crossing server or mobile gateway was disconnected for a short time. Handovers and other brief disconnections in the data connection between the train unit and level crossing server or mobile gateway were detected by looking at the delay with which the packets were received at the level crossing server (interface A, Figure 3). Log files of the utilization of rolling stock such as rail buses and diesel engines were obtained from the railway operator and matched to observed trains and lists of diesel engines and rail buses equipped with a train unit connected to the system. Twelve unequipped trains mentioned in the log files could be matched with high probability to trains observed at the level crossing but not detected by the system. However, an improved matching procedure is needed to get a more accurate estimate of the number of situations in which the cause of a missed alarm was an unequipped train.

Other identified causes of completely missed alarms were failures of some train units and temporarily occurring situations in which the in-vehicle system was inoperative. It was also suspected that some unsuccessful alarms were related to excessive delay to packets caused by the mobile gateway and the web service interface between the mobile gateway and the level crossing server. Among other factors affecting the reliability of the system, limitations of the GPRS data connection and the mobile platform used as an in-vehicle device were found to be significant sources of errors and randomness in the system.

The results obtained within a short monitoring period with a limited number of various types of events cannot be used to draw a conclusion on the impact of events that occur only rarely. Thus, the impact of server failure during the early days of the observation period had to be excluded, and the results obtained do not reflect the impact of the reliability of the level crossing server on the reliability of the system experienced by the end-user.

**Recommendations for further research**

More detailed analysis of the causes of false negatives and false positives would be required to better understand the behaviour of the system and to estimate the potential for improvement. The analysis of false negatives and false positives should make a clear distinction between faults likely to occur during normal operation of the system and causes related to the organisation of the field test such as rail buses or diesel engines not being equipped with the train unit connected to the system.

While relatively few problems could be attributed to the accuracy or performance of satellite positioning, the handovers related to the GSM network and long round trip times related to GPRS seemed to have a significant impact on the reliability of the system. Therefore, two main questions related to data communication and reliability of the system can be formulated.
The first important question is whether the limitations of existing GPRS data communication can be overcome with improvements to communication protocols and data processing algorithms in train units, level crossing server and the in-vehicle system and how much they will improve the reliability of the system. For example, handovers from one cell of the GSM network to another are likely to occur in the same place and several types of information can be used to predict the movements of the train during a handover procedure causing delay or packet loss for the messages sent by the train unit. Different techniques such as matching the observed train to timetables and historical speed profiles of similar locomotives or rail buses in time and space on the same railway line etc. should be tested. In cases where two separate GSM networks are available, two train devices with SIM cards from different network operators could be used in the same engine or rail bus to achieve some redundancy for the GPRS data connection. Using a transmission protocol other than TCP would also shorten the time needed to recover from a disconnection of the data link between train unit and level crossing server or mobile gateway.

The second important question is what will be the impact of new data communication services offered by 3G and 4G networks on the reliability of the system. It is probable that shorter data latency and reduced packet loss will improve the operation of the system, but the impact on the system via handover performance is not fully predictable. Empirical tests and detailed analysis of handovers in UMTS and LTE and vertical handovers from UMTS or LTE to GPRS offered to 2G networks and vice versa would be needed to get an accurate picture of the impact of UMTS and LTE on the reliability of the system.

7. REFERENCES


Methods for assessing the pedestrian level of service: International experience and adjustment to the Greek walking environment - The case of Thessaloniki
1. **INTRODUCTION**

For some time, transportation engineers and planners have paid attention primarily to the motorized transportation system. Even today, the motorized transportation system receives an overwhelming priority over systems that serve the needs of non motorized users such as pedestrians and bicyclists. However, in recent years, emphasis has been placed on multimodal approaches in order for the challenges of congestion, air quality, infrastructure concurrency and quality of life to be met. Many communities started promoting walkability concepts through education and infrastructure improvements. The latter, created the necessity of measuring the performance of the facilities that serve pedestrian traffic in order to determine quality of operations, existing deficiencies, needs for improvements and priority setting. One of the most common approaches used to assess transportation facilities is the concept of Level Of Service (LOS).

This paper aims to present, analyze and evaluate the methods that have been developed worldwide for assessing the pedestrian LOS and subsequently investigate their applicability to the Greek urban walking environment. In particular, eighteen (18) methods categorized by the spatial level of pedestrians’ movements in an urban environment are analyzed. The primary factors that each method takes into consideration for assessing the pedestrian LOS as well as the characteristics of the walking environment, where the methods were developed, are described in order to assess their applicability and suitability under various circumstances and specifically for a case study in the city of Thessaloniki.

In the case study, five methods were selected and applied in order to evaluate the pedestrian LOS along the sidewalk of an arterial street in the city center. According to the study’s results, the estimated pedestrian LOS varies significantly depending on the selected method, as some take into consideration mainly qualitative parameters where others use only quantitative ones, and none of the considered methods by itself could accurately predict the actual conditions of the pedestrian movements.

2. **METHODS FOR ASSESSING THE PEDESTRIAN LEVEL OF SERVICE (LOS)**

The following sections present the eighteen (18) methods that have been developed over the last decade, in various countries, for assessing the pedestrian LOS, categorized by the spatial level of pedestrians’ movements in an urban environment.
2.1 ON AREAS

This category includes only one method, developed in the City of Kansas, USA as part of its Walkability Plan in order to promote and enhance walking as a viable form of transportation at the citywide level. According to this method, the pedestrian LOS measures, which are only qualitative ones, are as follows:

- ‘Directness’ (existence of sidewalks in the city area),
- ‘Continuity’ (completeness of the pedestrian system, determined by comparing the total length of sidewalks to the total length of streets within the area)
- ‘Street crossings’ (ability to safely cross a street based on the number and widths of lanes to cross
- ‘Visual interests & amenities’ (pedestrian’s systems attractiveness and appeal)
- ‘Security’ (pedestrian clearly visible to other pedestrians or activities).

It should be noted that the pedestrian LOS is estimated for each one of the above parameters separately, as the method does not provide the calculation of an overall pedestrian LOS (Kansas Walkability Plan, 2004)

2.2 ON ROADWAY SEGMENTS (WITHOUT CONSIDERING THE IMPACT OF INTERSECTIONS)

- **Jaskiewicz F.**

This method evaluates the pedestrian LOS on a roadway segment based on nine qualitative evaluation measures: (i) enclosure/definition, (ii) complexity of path network, (iii) building articulation, (iv) complexity of spaces, (v) overhangs/awnings/varied roof lines, (vi) buffer, (vii) shade trees, (viii) transparency and (ix) physical components/condition. A simple rating scale of 1 (=very poor) to 5 (=excellent) is applied, to assess the degree to which certain target areas conform to the nine evaluation measures. The scores are then aggregated and averaged to obtain an overall pedestrian LOS (A-F). (Jaskiewicz, 2000)

- **Gallin N.**

This method is used for assigning a LOS grade on a roadway segment, based on eleven (11) factors classified as follows: design (path width, surface quality, obstructions, crossing opportunities, support facilities), location (connectivity, path environment, potential for vehicle conflict) and users (pedestrian volume, mix of path users, personal security). For each of these factors a score is given (0-4) as it relates to the roadway segment under investigation. For example, when assessing the width of a path, the score allocated will be 0 if there is no pedestrian path and 4 if the path is more than 2m wide. After multiplying each score with its relative weighting, the total weighted score (for all the 11 factors) is then cross-referenced with a table to determine the actual pedestrian LOS (A-E). (Gallin, 2001)
- **Landis B. et al**

The research team developed a mathematical model for assessing the pedestrian LOS using as primary independent variables: (i) the existence of sidewalk, (ii) the lateral separation elements between pedestrians and motor vehicle traffic, (iii) the motor vehicle traffic volume and (iv) the motorized vehicle speed. These variables were generated through a field study, where the participants’ reactions during a walking course were recorded, regarding their sense of safety or comfort within the examined roadway segment. The model’s numerical result is then cross-referenced with a table to determine the pedestrian LOS (A-F). (Landis et al, 2001)

- **Florida Department of Transportation (FDOT)**

For assessing pedestrian LOS, the FDOT developed a model that was based on the same study as the model of Landis et al. For that reason, the two models take into consideration the same variables and the only difference is concentrated on the constant term and the variables’ coefficients (Q/LOS Handbook, 2002)

- **Mozer D.**

According to this method, the suitability of roadway segments for pedestrians is based on four (4) primary variables: walkarea width-volume, walkarea-outside lane buffer, outside lane traffic volume, outside lane motor vehicle speed, plus three secondary factors: walkarea penetrations, heavy vehicle volumes and intersection waiting time. For each primary variable, a “stress level” (SL) is measured and set on a scale of 1 to 5. The secondary factors are added as decimals to the primary SL. The higher the SL, the lower the pedestrian LOS. The SLs are added and then averaged to determine the overall pedestrian LOS (A-E). It should be noted that evaluations are made mid-block on a discrete segment of the facility. (Mozer, 1997)

- **Jensen S.**

The Jensen’s method was developed in Denmark and was based on a field study for objectively quantifying pedestrian stated satisfaction with road sections between intersections. The findings of the study resulted in developing a utility function taking into consideration: the type of the walking area, the type of roadside development or landscape, the motor vehicles per hour in both directions, the average motor vehicle speed (km/h), the passed pedestrians per hour on nearest roadside, the bicycles and mopeds per hour in both directions, the width of the buffer (m), the parked motor vehicle per 100m, the width of walking area, the total width of walking area and nearest drive lane and median, drive lane and tree dummies. Though this utility function, six LOS designations are then defined (A-F). (Jensen, 2007)
2.3 On roadway corridors and arterials

- Dixon L.

Dixon’s method was developed and implemented through the Gainesville Mobility Plan, which was used as a congestion management system plan for the city of Gainesville, Florida. The pedestrian LOS performance measures evaluate roadway corridors using a point system of 1 to 21 that results in LOS ratings from A to F. The point system takes into consideration: (i) the pedestrian facility provided, (ii) conflicts, (iii) amenities, (iv) motor vehicle LOS, (v) maintenance and (vi) transportation demand management (TDM) programs or multimodal links to transit. The, under investigation, roadway corridor is divided in roadway segments, which are evaluated based on the above parameters. The segments’ scores are then multiplied with their weight and the corridor’s score is calculated as a sum of the adjusted segments’ scores. Pedestrian LOS ratings are then defined using the corridor score. (Dixon, 1996)

- Landis B. et al

The research team developed a LOS model that represents pedestrians’ perceptions of how well urban arterials with sidewalks meet their needs. Based on a field study, conducted by the researchers, it was revealed that traffic volumes on the adjacent roadway and the density of conflict points along the facility are the primary factors in the model for pedestrians traveling along urban arterial with sidewalks. In order to define the pedestrian LOS (A-F), the model’s numerical result is cross-referenced with a table. (Landis et al, 2006)


The primary performance measure for assessing the pedestrian LOS, according to the HCM method, is the average pedestrian walking speed (m/s), which is determined by the average time that is required for a pedestrian to cross a certain distance, including delays. Once determined, the average walking speed is cross-referenced with a table to define the pedestrian LOS (A-F). (HCM, 2000)

2.4 On intersections (not considering only crosswalks)

The city of Charlotte, North Carolina developed a method to assess the design features that affect pedestrians on signalized intersections, focusing not only on crosswalks but examining the whole intersection area. Among the key features identified and rated are: (i) crossing distance, (ii) signal phasing & timing, (iii) corner radius, (iv) right-turns-on-red, (v) crosswalk and (vi) traffic flow direction. After been identified, those key elements are relative weighted using a point system. Points are assigned according to how well they achieve these objectives.
Adding points from all the above six parameters for a particular intersection yields the pedestrian LOS (A-F) for a given intersection. (Steinman, Hines, 2003)

2.5 ON PEDESTRIAN CROSSINGS (FOR MID-BLOCK AND INTERSECTIONS)

- **Highway Capacity Manual 2000**

The method described in the HCM 2000 enables the assessment of the pedestrian LOS specifically on pedestrian crossings at both signalized and unsignalized intersections. Regarding signalized intersections, two approaches are described. In the first approach, as primary factor to assess the pedestrian LOS is considered to be the average pedestrian delay (s), while the second approach considers the circulation area per pedestrian (m²/p). In unsignalized intersections, the average delay/pedestrian (s), which is calculated based on the vehicular flow rate (veh/s) and the group critical gap, is used in order to determine the pedestrian LOS (A-F). (HCM, 2000)

- **Landis B. et al**

The research team developed a pedestrian LOS model that accurately represents pedestrians’ perceptions of crossing at signalized intersections. Based on the findings of a study, conducted by the research team, the primary factors that are taken into consideration in the model are: (i) right-turn-on-red volumes for the street being crossed, (ii) permissive left turns from the street parallel to the crosswalk, (iii) motor vehicle volume on the street being crossed, (iv) midblock 85 percentile speed of the vehicles on the street being crossed, (v) the number of lanes being crossed, (vi) the pedestrians’ delays and (vii) the presence or absence of right-turn channelization islands. The numerical result of the model is then cross-referenced with a table in order to determine the pedestrian LOS. (Landis et al, 2005)

- **Chu X. and Baltes M.**

This method was developed in U.S.A. and its model was based on a large number of data, collected through a field survey. The determinants that were used in the model were: (i) the age of the pedestrians, (ii) the traffic volume, (iii) turning movements, (iv) the traffic speed, (v) the crossing distance, (vi) restrictive and non-restrictive medians, (vii) crosswalks, (viii) pedestrian signals, (ix) signal cycle and (x) signal spacing. (Baltes, Chu, 2001)

- **Muraleetharan et al**

This method was developed in Japan and considered as primary factors for assessing the pedestrian LOS on a crosswalk: (i) the level of space at corners, (ii) the crossing facilities, (iii)
the turning vehicles and (iv) the pedestrian delay. It is based on the concept of total utility value, which comes from a conjoint analysis research. To estimate the total utility value for pedestrian crossings, field measurements must be carried out to collect operational and geometrical characteristics of the pedestrian crossings regarding the above factors. Calculated total utility values are then used to assign an overall LOS designation (A-F) to each crossing based on the assumption of a linear relationship between them. (Muraleetharan et al, 2004)

2.6 ON SIDEWALKS AND OTHER PEDESTRIAN FACILITIES

- *Highway Capacity Manual 2000*

The method described in the HCM 2000, evaluates the pedestrian LOS on uninterrupted-flow facilities, including exclusive pedestrian paths, queuing areas and shared pedestrian-bicycle facilities. The primary performance measure for walkways and sidewalks is space. However, for simplicity of field observation, pedestrian unit flow rate (p/min/m) is used as a service measure. The average space available to pedestrians can also apply for queuing areas. On shared facilities, the pedestrian LOS depends on the frequency that the average pedestrian is overtaken by bicyclists. (HCM 2000)

- *Muraleetharan T. et al*

Following the method that was developed for pedestrian crossings, this method is used for the assessment of sidewalks based on the concept of total utility value. The factors used to determine the total utility value are: (i) the lateral separation of the pedestrians, (ii) the width of sidewalks, (iii) obstructions, (iv) pedestrian flow rate and (v) number of bicycle passing and opposing events. (Muraleetharan et al, 2004)

- *Tan D. et al*

According to the experience nation-wide and abroad as well as the actual traffic condition of urban roads in China, the research team developed a LOS model, including as primary factors: (i) the road transect form, (ii) the pedestrian flow characteristics, (iii) the vehicle and bicycle flow characteristics, (iv) the obstructions and (v) the frequency of the driveway access. The numerical result of the model is then cross-referenced with a table in order to determine the pedestrian LOS. (Tan et al, 2007)

The following table summarizes the methods described above based on the spatial level they refer to and the primary factors that take into consideration for assessing the pedestrian LOS. These factors are categorized as design, traffic, environmental and delay ones.
Table 1. Methods for assessing the pedestrian LOS based on the spatial level they refer to and the primary factors they consider (Lopez, 2006; Sdoukopoulos, 2009)

<table>
<thead>
<tr>
<th>Method</th>
<th>Country</th>
<th>Present</th>
<th>Spatial level</th>
<th>Primary factors</th>
<th>Delays</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kansas Walkability Plan (2004)</td>
<td>U.S.A.</td>
<td>✔️</td>
<td>Area</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>FDOT (2002)</td>
<td>U.S.A.</td>
<td>✔️</td>
<td>Design</td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td>HCM 2000 (TRB)</td>
<td>U.S.A.</td>
<td>✔️</td>
<td>Delays</td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td>Chu X., Baltes M. (2001)</td>
<td>U.S.A.</td>
<td>✔️</td>
<td></td>
<td></td>
<td>✔️</td>
</tr>
</tbody>
</table>
3. **Comparison and Evaluation of Pedestrian Level of Service (LOS) Methods**

In this chapter an attempt is made to compare and evaluate the LOS methods, described above, based on the primary factors they consider. The methods follow the same categorization, as previously, excluding from the following analysis the ones that evaluate the pedestrian LOS on areas and intersections as no other method is included in those categories. It should be noted that for each category different factors are used for the evaluation, based on the characteristics of the study area.

3.1 **On roadway segments (without considering the impact of intersections)**

The following table summarizes the six (6) methods that can be applied to assess the pedestrian LOS on roadway segments and the primary factors they consider.

Table 2. Comparison of the methods that can be applied to assess the pedestrian LOS on roadway segments based on their primary factors

<table>
<thead>
<tr>
<th>Method</th>
<th>Volume</th>
<th>Traffic conflicts</th>
<th>Safety / Comfort</th>
<th>Users’ perceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jaskiewicz F.</td>
<td>❌</td>
<td>❌</td>
<td>✓ (Buffer, transparency)</td>
<td>❌</td>
</tr>
<tr>
<td>Gallin N.</td>
<td>✓ (Pedestrian volume, mix of path users)</td>
<td>✓ (Potential for vehicle conflict per km, crossing opportunities)</td>
<td>✓ (Personal security)</td>
<td>❌</td>
</tr>
<tr>
<td>Landis B. et al</td>
<td>✓ (Vehicle volume and speed)</td>
<td>❌</td>
<td>✓ (Sidewalk, buffer and on-street parking)</td>
<td>✓</td>
</tr>
<tr>
<td>Mozer D.</td>
<td>✓ (Vehicle volume and speed, pedestrian volume)</td>
<td>❌</td>
<td>✓ (Buffer)</td>
<td>❌</td>
</tr>
<tr>
<td>Jensen S.</td>
<td>✓ (Vehicle, pedestrian and bicycle volume)</td>
<td>❌</td>
<td>✓ (Buffer, type of roadside development, trees)</td>
<td>❌</td>
</tr>
<tr>
<td>FDOT</td>
<td>✓ (Vehicle volume and speed)</td>
<td>❌</td>
<td>✓ (Sidewalk, buffer and on-street parking)</td>
<td>❌</td>
</tr>
</tbody>
</table>

According to the above table, vehicle volume is considered as an important factor as most of the examined methods take it into consideration, fact that highlights its large impact on pedestrian safety and comfort, in comparison to pedestrian volume, which proves to be of less importance. Traffic conflicts are also not greatly considered, as the methods don’t take into account the impact of intersections. Furthermore, the issue of safety and comfort receives a
different approach by each method, as various factors are considered. Finally, particular emphasis should be given to the fact that users’ perceptions were relatively considered by four of the six methods as a major component in their development.

3.2 **ON ROADWAY CORRIDORS AND ARTERIALS**

The following table summarizes the three (3) methods that can be applied to assess the pedestrian LOS on roadway segments and the primary factors they consider.

Table 3. Comparison of the methods that can be applied to assess the pedestrian LOS on roadway segments based on their primary factors

<table>
<thead>
<tr>
<th>Method</th>
<th>Volume</th>
<th>Traffic conflicts</th>
<th>Safety / Comfort</th>
<th>Users’ perception</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dixon L.</td>
<td>✓ (Vehicle volume)</td>
<td>X</td>
<td>✓ (Pedestrian facility provided, amenities, medians present)</td>
<td>X</td>
</tr>
<tr>
<td>Landis B. et al</td>
<td>✓ (Vehicle volume)</td>
<td>✓ (Density of conflicts points along the facility)</td>
<td>✓ (Lateral separation elements between pedestrians and vehicle volume)</td>
<td>X</td>
</tr>
<tr>
<td>HCM 2000</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

The methods under this category vary significantly regarding the factors they consider for assessing the pedestrian LOS. Specifically, vehicle volume is not considered by HCM, which however includes pedestrian speed and delays as primary factors. It should be noted that Dixon’s method focuses mainly on factors regarding pedestrian safety and comfort as it is mostly used as a design rather than evaluation tool. Finally, a common characteristic between the above methods is the fact that they were all developed in USA, with the HCM method to be considered as the one most widely used.

3.3 **ON PEDESTRIAN CROSSINGS**

The following table summarizes the four (4) methods that can be applied to assess the pedestrian LOS on roadway segments and the primary factors they consider.

Table 4. Comparison of the methods that can be applied to assess the pedestrian LOS on pedestrian crossings based on their primary factors
The majority of the above methods refer to pedestrian crossings on signalized and unsignalized intersections, with the only exception of the Chu and Baltes method, which determines the pedestrian LOS for mid-block crossings. As most significant factors, for assessing the pedestrian LOS were proven to be vehicle volume, pedestrian delay and turning vehicles. Finally, regarding crossing facilities, medians and ramps are the mostly considered elements.

### 3.4 ON SIDEWALKS AND OTHER PEDESTRIAN FACILITIES

The following table summarizes the three (3) methods that can be applied to assess the pedestrian LOS on sidewalks and other pedestrian facilities and the primary factors they consider.

<table>
<thead>
<tr>
<th>Method</th>
<th>Pedestrian crossing</th>
<th>Signalized – unsignalized intersections</th>
<th>Vehicle volume</th>
<th>Pedestrian delay</th>
<th>Turning vehicle</th>
<th>Crossing facilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCM 2000</td>
<td>On intersection</td>
<td>Both</td>
<td>✅</td>
<td>✅</td>
<td>✅ (signalized)</td>
<td>X</td>
</tr>
<tr>
<td>Landis B. et al</td>
<td>On intersection</td>
<td>Unsignalized</td>
<td>✅</td>
<td>✅</td>
<td>👎</td>
<td>Medians</td>
</tr>
<tr>
<td>Chu X., Baltes M.</td>
<td>Mid-block</td>
<td>X</td>
<td>✅</td>
<td>X</td>
<td>✅</td>
<td>Medians</td>
</tr>
<tr>
<td>Muraleetharan et al</td>
<td>On intersection</td>
<td>Both</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td>Medians, ramps</td>
</tr>
</tbody>
</table>

The above table indicates pedestrian volume as an important factor for assessing the pedestrian LOS on sidewalks and other facilities. However, the method, developed in China (Tan D.et al), also takes into consideration vehicle and bicycle volume, as it considers that those factors have a great impact on pedestrian safety. Further elements affecting pedestrian safety proved to be obstructions and the lateral separation between pedestrians and vehicles.
4. **CASE STUDY IN THE CITY OF THESSALONIKI**

4.1 **CASE STUDY SETUP**

In order to investigate both the applicability and the suitability of some of the above methods in the Greek urban walking environment, a case study was performed in the city of Thessaloniki. Five appropriate methods were selected and applied in order to evaluate the pedestrian LOS along the sidewalk of an arterial street in the city centre (Figures 1 & 2). The arterial street was a three-lane street and in the selected sidewalk, which was exclusively used by pedestrians, a large number of cafes and shops were located.

![Figure 1: The arterial street in the centre of Thessaloniki](image1)

![Figure 2: The selected sidewalk](image2)

4.2 **SELECTED METHODS**

As the study area is a sidewalk on roadway segment between two consecutive intersections, the methods that were selected and applied, based on previous categorization were the Jaskiewicz and Gallin methods, which relate to roadway segments and the HCM 2000, Muraleetharan T. et al and Tan D. et al methods which relate to sidewalks. The selection of the methods was based mainly on the scale of the study and the fact that those methods take into consideration, for assessing the pedestrian LOS, parameters that can be easily measured both qualitative and quantitative ones.

In order to collect the necessary data, after defining the qualitative and geometrical characteristics of the study area, measurements regarding the pedestrian and vehicle volume were conducted in 15min time periods, both in peak and off-peak hours.
4.3 Calculating the Pedestrian LOS

- Jaskiewicz F.

The roadway segment was evaluated through nine qualitative measures, using a simple rating scale from 1 (=very poor) to 5 (=excellent). The score for each one of the nine measures and its justification comments are presented in the following table.

Table 6. Rating and comments regarding the parameters that are used for assessing the pedestrian LOS according to Jaskiewicz’s method

<table>
<thead>
<tr>
<th>Measures</th>
<th>Score</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enclosure/definition</td>
<td>5</td>
<td>Building constructed side-by-side along the sidewalk ( \Rightarrow ) edges of the street well defined</td>
</tr>
<tr>
<td>Complexity of path network</td>
<td>3,5</td>
<td>The path network furnishes pedestrians with numerous route choices</td>
</tr>
<tr>
<td>Building articulation</td>
<td>2</td>
<td>Poorly articulated buildings</td>
</tr>
<tr>
<td>Complexity of spaces</td>
<td>3</td>
<td>Quite frequent variation in the orientation and character of public spaces</td>
</tr>
<tr>
<td>Overhangs/awnings/varied roof lines</td>
<td>4</td>
<td>Large number of overhangs and awnings improving the comfort level of pedestrians</td>
</tr>
<tr>
<td>Buffer</td>
<td>1</td>
<td>Poorly buffered street</td>
</tr>
<tr>
<td>Shade trees</td>
<td>3</td>
<td>Large number of shade trees</td>
</tr>
<tr>
<td>Transparency</td>
<td>4</td>
<td>Transparency is created through the use of windows, outdoor displays and sidewalks cafes</td>
</tr>
<tr>
<td>Physical components/condition</td>
<td>3,5</td>
<td>Acceptable structural integrity and functionality of the sidewalk</td>
</tr>
<tr>
<td><strong>Average score</strong></td>
<td><strong>3,14</strong></td>
<td><strong>Level Of Service: C</strong></td>
</tr>
</tbody>
</table>

- Gallin N.

The roadway segment was evaluated through eleven factors categorized as design, location and user ones. The criteria for each factor, determined the scores (0-4), which are presented in the following table.

Table 7. Rating and criteria regarding the factors that are used for assessing the pedestrian LOS according to Gallin’s method

<table>
<thead>
<tr>
<th>Category</th>
<th>Factor</th>
<th>Score</th>
<th>Weight</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Design factors</strong></td>
<td>Path width</td>
<td>2</td>
<td>4</td>
<td>1,1 – 1,5 m</td>
</tr>
<tr>
<td></td>
<td>Surface quality</td>
<td>3</td>
<td>5</td>
<td>reasonable quality, i.e. acceptable standard</td>
</tr>
<tr>
<td></td>
<td>Obstructions</td>
<td>0</td>
<td>3</td>
<td>more than 21 obstructions per km</td>
</tr>
<tr>
<td></td>
<td>Crossing opportunities</td>
<td>2</td>
<td>4</td>
<td>some provided and are reasonably well located but more are needed</td>
</tr>
<tr>
<td></td>
<td>Support facilities</td>
<td>0</td>
<td>2</td>
<td>non existent</td>
</tr>
<tr>
<td><strong>Location factors</strong></td>
<td>Connectivity</td>
<td>3</td>
<td>4</td>
<td>good</td>
</tr>
<tr>
<td></td>
<td>Path environment</td>
<td>1</td>
<td>2</td>
<td>poor environment, may be within 1m of kerb</td>
</tr>
<tr>
<td></td>
<td>Potential for vehicle conflict (per km)</td>
<td>3</td>
<td>3</td>
<td>reasonable, 1 to 10 or less conflict points per km</td>
</tr>
</tbody>
</table>
After multiplying each score with its relative weighting, the total weighted score was used to determine the pedestrian LOS (A-E), which in this case was C.

- **HCM 2000**

After measuring the pedestrian flow rate (15min) and the effective walkway width, the pedestrian LOS was calculated (using appropriate tables), which for peak hours was B and for off-peak hours was A.

Table 8. Rating and criteria regarding the factors that are used for assessing the pedestrian LOS according to HCM 2000

<table>
<thead>
<tr>
<th>Hour</th>
<th>Pedestrian flow rate (ped/min/m)</th>
<th>Criteria</th>
<th>Pedestrian LOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak</td>
<td>6.38</td>
<td>&lt; 16</td>
<td>A</td>
</tr>
<tr>
<td>Off-peak</td>
<td>18.24</td>
<td>16-23</td>
<td>B</td>
</tr>
</tbody>
</table>

- **Muraleetharan T. et al**

The pedestrian LOS according to this method is estimated through four (4) parameters. Based on certain criteria, a utility value is determined for each parameter, which is then translated into pedestrian LOS (1-3).

Table 9. Rating and criteria regarding the factors that are used for assessing the pedestrian LOS according to Muraleetharan’s et al method

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Utility value</th>
<th>Level Of Service</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Width of sidewalk and lateral separation of the pedestrian</td>
<td>0,15</td>
<td>2</td>
<td>Width 1.5 - 3.0m and acceptable lateral separation of the pedestrians</td>
</tr>
<tr>
<td>Obstructions</td>
<td>-0.77</td>
<td>3</td>
<td>More than 5 obstructions per 100m</td>
</tr>
<tr>
<td>Pedestrian flow rate</td>
<td>1.53</td>
<td>1</td>
<td>Less than 24 ped /min/m in peak and off-peak period</td>
</tr>
<tr>
<td>Number of bicycle passing and opposing events</td>
<td>1.72</td>
<td>1</td>
<td>Less than 60 events per hour</td>
</tr>
<tr>
<td><strong>Total utility value</strong></td>
<td><strong>2.63 + constant term 3.90 = 6.53</strong></td>
<td><strong>1 (A-B)</strong></td>
<td><strong>This method uses three Levels Of service (1-3), in contradiction to the six levels (A-F) commonly used</strong></td>
</tr>
</tbody>
</table>
For applying this method, the following parameters were measured: vehicle volume of the outside lane of the adjacent road \( (Q_V) \), pedestrian volume \( (Q_P) \) and bicycle volume \( (Q_B) \) for 5min period, frequency of driveway access per m \( (P) \) and the distance between the sidewalk and the outside lane of the adjacent road \( (W_r) \). These measures were incorporated into the mathematical model: Ped LOS = \(-1.43 + \left(0.006 \times Q_B \right) – \left(0.003 \times Q_P \right) + \left(0.056 \times \frac{Q_V}{W_r} \right) + \left(11.24 \times (P – 1.17 \times P^3) \right)\). This value, according to the method, corresponds to pedestrian LOS A.

### 4.4 RESULTS OF THE CASE STUDY

The following table presents the pedestrian LOS for the study area, as calculated using the five (5) selected methods.

Table 10. Pedestrian LOS of the study area per method applied

<table>
<thead>
<tr>
<th>Methods</th>
<th>Roadway segment</th>
<th>Sidewalks and other pedestrian facilities</th>
<th>Jaskiewicz</th>
<th>Gallin</th>
<th>off peak</th>
<th>peak B</th>
<th>Muraleetharan T. et al</th>
<th>Tan D. et al</th>
</tr>
</thead>
</table>
| Level of Service (LOS) | C               | C                                       | A          | A-B       | A-

The above table indicates that the estimated pedestrian LOS varies depending on the selected method. Specifically, the methods provided by Jaskiewicz and Gallin result in the lowest pedestrian LOS, as these methods relate to roadway segments and consider mainly qualitative parameters ignoring traffic variations. Especially, the first one does not consider at all, pedestrian or vehicle volumes while the latter takes into consideration only pedestrian volume. This fact indicates that qualitative parameters seem stricter and depend largely on the evaluator’s judgment.

In comparison to those two methods, the one described in HCM does not consider the conditions of the walking environment and is based only on pedestrian flow rate, factor that can be misleading regarding sidewalks’ conditions, as sidewalks poorly maintained and not used by pedestrians receive, due to low pedestrian flow rate, a high pedestrian LOS.

Finally, the two methods described by Muraleetharan T. et al and Tan D. et al, use a combination of qualitative and quantitative parameters, resulting in more reliable pedestrian LOS.
5. **CONCLUSIONS**

This paper attempted to describe and evaluate eighteen (18) methods that have been developed worldwide for assessing the pedestrian LOS on different spatial levels, from a wide area to a specific pedestrian facility. A large variety of factors (qualitative and quantitative) are used within the different methods and in many cases different approaches are considered for measuring each factor. The majority of the examined methods were developed in U.S.A, fact that indicates that the produced LOS models can be best applied in the American walking environment. Although the use of these methods in other countries has a high validity, many countries acknowledged the need to develop a method that incorporates national walking characteristics, resulting in more reliable pedestrian LOS.

In Greece, the most commonly used method is the one described in the HCM 2000, as it can be easily applied and does not require complex data. However, this method cannot accurately predict the actual conditions of the pedestrians’ movements. According to the case study’s results, the estimated pedestrian LOS varies, depending on the selected method. Specifically, the methods that take into consideration mainly qualitative parameters result in lower pedestrian LOS that those that consider mainly quantitative ones.

Finally, by examining the findings of the case study in relevance to the actual walking conditions, interesting conclusions were drawn. The main conclusion was that there is a need to develop a new model to assess the pedestrian LOS in Greece through the study and selection of those parameters, used in the previously described methods that are considered to describe well the Greek walking environment, which is characterized by sidewalks with small widths and many obstructions, poor personal security (no lateral separation of pedestrians and vehicles) and many potential conflicts with the vehicles, which dominate at a large extent the urban environment. This paper can provide the theoretical background for the development of such a model, taking into consideration those specific characteristics of the Greek walking environment and documenting the perceptions of the Greek pedestrians, regarding the factors that they consider as the most significant ones in a walking environment.

6. **REFERENCES**


Quality/Level of Service Handbook 2002, State of Florida, Department of Transportation


Optimization of key design elements of 2+1-routes

- Final paper -
1. INTRODUCTION, SCOPE AND AIMS

A 2+1-route is characterised by a single carriageway with a three-lane cross section. It consists of two-lane sections with a passing lane and one-lane sections (also called feeder sections), positioned before, behind and beside each passing section (Figure 1).

![Figure 1: Schematic representation of operational form of 2+1-routes (Priemer, 2004).](image)

While traffic in a passing section has the possibility to pass slower vehicles, passing is not allowed for traffic travelling in the opposite direction. This causes the traffic in the opposite direction to ‘bunch up’, or form platoons within the one-lane sections. To avoid this, and ensure that the negative effects are minimised, passing lanes should be provided in an alternating manner, in each direction, along the whole 2+1-route. Furthermore, if vehicle platoons are not dissolved at the end of the passing sections, the level of service is impaired, and negative impacts on traffic safety will be the likely result.

In recent years a number of research projects have dealt with these issues (predominantly traffic safety and traffic flow) for 2+1-routes (Weber & Löhne, 2003, Brannolte et al., 2004, Gattis et al., 2006). However, up to now there has been no comprehensive finding (other than model-based considerations) regarding the passing process, and the dissolving of platoons as dependant on marginal conditions relating to design and operation. Therefore, the primary aim of the dissertation (Irzik 2009) – that forms the basis of this paper – was to develop a procedure in order to determine the optimal length of passing sections on 2+1-routes, on the basis of empirical studies and results.

The optimization of passing section length for 2+1 routes in order to dissolve the platoons is difficult, as there is a conflict between design objectives. With regard to traffic safety passing sections must be long enough so that all entering platoons are dissolved at the termination of the passing section. If a driver is not able to pass the platoon leader and therefore the platoon is not dissolved, the driver is impeded in his wish to drive as fast as he likes (under consideration of the environment and speed limit). This leads to frustration, which leads to aggressiveness and at the end to risky passing maneuvers. However, passing sections must not be too long, as this would have an adverse effect on the length of the opposing one-lane section and thus on platoon formation. Apart from determining the dimension of passing sections based on the dissolving of platoons, aspects of traffic safety were also investigated. A
research project investigating the passing process on 2+1 routes (Friedrich et al., 2005), commissioned by the German federal highway research institute (BASt), constitutes the basis of the presented dissertation (Irzik 2009).

2. METHODOLOGY

Summarising the current scientific knowledge concerning this topic was the first stage of the research (see Figure 2). After a description of the choice of investigated sections, the relevant parameters of traffic flow, passing process, platooning, and dissolving of platoons were introduced. Following the description of the developed and applied investigation methodology for the empirical investigations, the procedure of the analysis of empirically obtained data, as well as the used methods/techniques was listed. As a rule, the obtained results were represented statistically, and if possible compared to results of previous studies. Initially, the speed and passing process (particularly at the beginning and the end of passing) and platooning, both ahead of, and within 2+1 routes were analysed. Afterwards different approaches to the determination of the length of passing sections required for dissolving platoons on 2+1-routes were tested. Finally, some recommendations were provided due to the gained new findings.

Figure 2: Research process steps.
3. ANALYSIS OF LITERATURE

All studies dealing with traffic safety on 2+1-routes during recent years show that 2+1-routes are characterised by a level of higher traffic safety, expressed in terms of accident-cost rate, in comparison to other single carriageway cross sections outside built-up areas. However, due to the comparatively higher level of travel speeds, the average accident severity is slightly higher in comparison to narrower two-lane rural roads.

Weber & Löhe (2003) could not confirm safety concerns regarding the opening of 2+1-routes for mixed traffic. While they reported low accident figures for these types of roads with mixed traffic, this was based on observations of roads with only a marginal share of bicycle and agricultural vehicle traffic, so firm conclusions regarding the safety concerns could not be drawn. Furthermore, simulation results obtained under the development of a design procedure for the update of the German Highway Capacity Manual (HBS) (FGSV, 2001) proved an explicit negative influence of slow moving traffic on the quality of traffic flow (Baselau, 2006, Brannolte et al., 2004).

![Figure 3: Selected junction types according to the new guidelines for the design of rural roads (RAL) currently being developed (FGSV, 2008).](image)

The work of Weber & Löhe (2003) provides some advice on the minimum length of passing sections for safety reasons. A minimum length of 1,000 m for passing lanes within 2+1 routes with mixed traffic was recommended. This finding was based on the results of their accident analysis, which showed a disproportionally high number of accidents on sections shorter than 1,000 m compared to sections with lengths exceeding 1,000 m. It was also found that an adapted alignment, in combination with partially grade-separated junctions (Figure 3) not only has a positive influence on the level of homogeneous speed, but also on a lower accident-cost rate.

Palm & Schmidt (1999) had already stated that (in a similar manner as for accidents on sections between junctions) the accident rate and accident-cost rate for accidents at junctions are lower on 2+1 routes than on single carriageway two-lane cross sections. This can most likely be explained by the fact that the number of at-grade junctions on 2+1-routes is smaller than on two-lane rural road cross sections. The 2+1-route on the Federal Road B10 near Landau in Germany investigated by Kölle (1999) contained one partially grade separated and one partially at-grade junction (Figure 3). Despite a speed limit of 70 km/h in the area of the partially at-grade junction, the accident-cost rate reported was more than twice that of the partially grade separated junction. Considering that the accident rate was almost equal, this
result indicates a substantially higher level of accident severity in the partially at-grade junction. This finding was determined to be a result of the higher speeds within the junction area, which were observed to be higher than the speed limit. Hence, the predominant number of accidents at the partially at-grade junction occurred in the area of the junction with the B10 major road.

In contrast, at the partially grade-separated junction, only few accidents were recorded. These were mainly ‘damage only’ type accidents that arose in situations involving vehicles changing lanes in the merging area. Main accident locations were the at-grade parts of these partially grade separated junctions at the minor road. Kölle (1999) formed a ‘composite systems’ approach in order to determine the influence of junctions on the adjoining sections. These composite systems comprised the junction itself, and 500 m of the adjoining sections. Therefore, composite system No. 1 includes the partially at-grade junction, and No. 2 the partially grade-separated junction. The comparison between these systems showed that the accident-cost rate of system No. 2 was approximately one-fifth of the accident-cost rate determined for the system No. 1.

Kölle’s investigation included two further junctions; a partially grade-separated junction and an at-grade junction, on a 2+1 route (the B49, in Germany). The accident-cost rate at this partially grade-separated junction was slightly higher than at the partially at-grade junction on the B10. However, it was only half of the rate of the at-grade junction along the same 2+1 route. At the partially grade-separated junction on the B49 there were also only a few accidents caused by entering or exiting vehicles. Similar to the partially grade-separated junction on the B10, the one on the B49 is characterised by a high level of safety, compared to other partially grade-separated junctions investigated by Kölle.

It was found that the existing findings regarding passing process and dissolving of platoons were lacking adequate empirical verification. Hence, the procedure developed by Roos (1989) is based on theoretical considerations. For this reason there were still uncertainties in the determination of optimal section lengths as dependant on traffic parameters. Therefore, it was a key requirement to increase the knowledge and understanding of passing processes, and dissolving of platoons.

The literature showed that differences exist between the definitions of platoons. Following previous research (Roos, 1989, Brannolte et al., 2004), in this investigation a vehicle was assigned to a platoon if the gross time gap to the vehicle ahead or the platoon leader was below 3 seconds. In this case the platoon leader's speed must not exceed 90 km/h. As a supplementary condition within this investigation, it was decided that a vehicle must pass the platoon leader in the passing section, otherwise it was assumed that vehicles which do not pass the platoon leader have already reached their desired speed in the feeder section and are therefore not obstructed in their movement. Using similar logic, it was decided that the platoon leader does not belong to the platoon. As a result of the chosen investigation methods,
the results of previously described studies could not be compared with those developed here, because of difference between platoon definitions and conditions.

4. CHOICE OF TWO-LANE SECTIONS

As determined from previous studies, the traffic situation on rural roads is influenced by a large number of factors. These could be related to the traffic itself, or the design of the road or operation and because of the unknown extent of the correlation that may be caused, it is often difficult to quantify the influence of individual factors to any selected factors. In order to obtain meaningful results, as many potential influencing factors as possible should be kept constant. Thus, only 2+1 routes with the characteristics listed below were considered in this investigation. They can be considered as being characteristic for 2+1 routes in Germany.

- operated as a road for motor vehicles only
- maximum speed of 100 km/h
- only partially grade-separated junctions are present
- comprising a slight gradient (±2 %)
- minimal or low horizontal curvature.

<table>
<thead>
<tr>
<th>Investigated section</th>
<th>BP = bypass</th>
<th>dt = direction towards</th>
<th>Federal State</th>
<th>Road No.</th>
<th>adt [veh/24]</th>
<th>HGV-share [%]</th>
<th>Position</th>
<th>Transition</th>
<th>Length [m]</th>
<th>Lmedia [m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 BP Jever, dt Wittmund</td>
<td>NI</td>
<td>B 210</td>
<td>13.000</td>
<td>8</td>
<td>1 lane add.</td>
<td>1.474</td>
<td>2.200</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 BP Jever, dt Wittmund</td>
<td>NI</td>
<td>B 210</td>
<td>13.000</td>
<td>8</td>
<td>2 lane add.</td>
<td>1.208</td>
<td>1.500</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 BP Jever, dt Wilhelmshaven</td>
<td>NI</td>
<td>B 210</td>
<td>13.000</td>
<td>8</td>
<td>2 lane add.</td>
<td>1.498</td>
<td>1.100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Kirchhain / Cölbe, dt Marburg</td>
<td>HE</td>
<td>B 62</td>
<td>13.000</td>
<td>8</td>
<td>1 lane add.</td>
<td>1.092</td>
<td>2.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Kirchhain / Cölbe, dt Marburg</td>
<td>HE</td>
<td>B 62</td>
<td>13.000</td>
<td>8</td>
<td>2 lane add.</td>
<td>1.687</td>
<td>1.400</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Kirchhain / Cölbe, dt Kirchhain</td>
<td>HE</td>
<td>B 62</td>
<td>13.000</td>
<td>8</td>
<td>1 lane add.</td>
<td>1.706</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Paderborn / Schlangen, dt Horn-Bad Meinberg</td>
<td>NW</td>
<td>B 1</td>
<td>17.000</td>
<td>10</td>
<td>3 lane add.</td>
<td>828</td>
<td>1.800</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Paderborn / Schlangen, dt Paderborn</td>
<td>NW</td>
<td>B 1</td>
<td>17.000</td>
<td>10</td>
<td>1 lane add.</td>
<td>1.403</td>
<td>3.500</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Paderborn / Schlangen, dt Paderborn</td>
<td>NW</td>
<td>B 1</td>
<td>17.000</td>
<td>10</td>
<td>3 lane add.</td>
<td>1.195</td>
<td>1.400</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Niederbiel / Leun, dt Limburg an der Lahn</td>
<td>HE</td>
<td>B 49</td>
<td>19.000</td>
<td>11</td>
<td>1 lane add.</td>
<td>895</td>
<td>2.600</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 BP Straubing, dt Landau an der Isar</td>
<td>BY</td>
<td>B 20</td>
<td>20.000</td>
<td>16</td>
<td>1 lane add.</td>
<td>1.296</td>
<td>1.200</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 Niederbiel / Leun, dt Limburg an der Lahn</td>
<td>HE</td>
<td>B 49</td>
<td>19.000</td>
<td>11</td>
<td>uncritical</td>
<td>929</td>
<td>1.200</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 BP Straubing, dt Landau an der Isar</td>
<td>BY</td>
<td>B 20</td>
<td>20.000</td>
<td>16</td>
<td>4 uncritical</td>
<td>1.146</td>
<td>1.400</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 BP Dachau, dt Dachau</td>
<td>BY</td>
<td>B 471</td>
<td>16.000</td>
<td>10</td>
<td>1 uncritical</td>
<td>1.258</td>
<td>1.500</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 BP Dachau, dt Fürstenfeldbruck</td>
<td>BY</td>
<td>B 471</td>
<td>16.000</td>
<td>10</td>
<td>2 uncritical</td>
<td>1.353</td>
<td>1.600</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Investigated sections
The fifteen sections investigated here are listed in Table 1. Eleven of these sections begin with an added lane (Figure 4), and the other four with an uncritical transition (Figure 5). The eleven sections with a lane addition cover a wide spectrum of section lengths, between approximately 800 m to 1,700 m. The four sections starting with an uncritical transition vary between approximately 900 m to 1,400 m in total length. Seven investigated passing sections are located at the beginning of the 2+1 route (Position 1) (Figure 1). Consequently, the seven relating feeder sections are located outside of the actual 2+1 route. Therefore, they are part of a “normal” two-lane cross section. In contrast, the remaining eight investigated passing sections (Position 2 or higher), as well as their feeder sections, are within the 2+1 routes. Each of these feeder sections is also an opposite direction lane of another passing section.

5. DETERMINATION OF OPTIMAL LENGTHS OF PASSING SECTIONS

In the scope of this study several approaches for the determination of required passing section lengths were examined. It is obvious that the length required for the passing process depends (amongst other things), on the platoon length which builds up in the feeder section. Figure 6 shows the number of observed platoon lengths.

From the literature, the 85 % percentile platoon length (with regard to an interval) is seen as the decisive parameter for the dimensioning of the passing sections (Figure 7). However, in the analysis of platooning it was found that this parameter cannot be determined logically and/or with a sufficiently accurate congruence on the basis of the data collected under this investigation.
A procedure for the determination of the section length on the basis of empirically monitored platoon dissolving distances (Figure 8) also shows some weak points. Such a procedure was also developed in the scope of this study, and was generally based on previous theoretical discussions (Roos, 1989). This procedure has the decisive disadvantage of not taking into account un-dissolved platoons.

Figure 6: Number of observed platoon lengths \( (n=3,163) \).

Figure 7: Number of 85% platoon length \( (n=111) \).

Figure 8: Platoon dissolving distance in dependency of the platoon length (Box-Whisker-Plots) \( (n=1,384) \).
Furthermore, in this procedure it is necessary to consider influencing factors, including:

- those which result from the design and lay-out of the transition at the beginning
- the (absolute) position or the length of the passing sections
- the traffic volume, share of HGVs and the probably a prohibition of overtaking for HGVs.

All these factors substantially increase the difficulty of this procedure. Separate from this, a different and much easier approach was pursued. In this approach, the rate of platoons that were not dissolved are taken as a level of service in the determination of the passing section length.

In the analysis of platoon dissolving, there were some platoons on individually investigated sections that could not be dissolved fully on the current lengths of those sections. Based on the total quantity of platoons during the investigation period, the rate of un-dissolved platoons could be obtained. This rate is particularly dependant on the traffic volume, and also on the section length. Thus, if a level for the rate of un-dissolved platoons is pre-set, the required section length can be determined as dependant on traffic volume, based on this correlation.

Influencing factors arising from the length or position of the overtaking sections are implied in this procedure. All but one section where platoons were not fully dissolved were found only on sections with a lane addition. Thus, it was determined that there was no need to explicitly provide for the way the platoon leader merges, in the form of a supplement to the length a platoon needs for dissolving. From the insights obtained into the platoon leaders' lane changing behaviour it appeared admissible to adapt the required section lengths for the dissolving process by a reduction, for the cases beginning with an uncritical transition. This is because on sections with an added lane, the platoon leaders generally merged after a distance of 125 m, while on sections that began with an uncritical transition, the first vehicle to pass the platoon leader merged at the latest after 50 m in 85% of all cases. Hence, the appropriate reduction was determined to be 75 m.

No useful results regarding the maximum platoon share which should be avoided for safety reasons can be gained from a comparison of parameters of platooning and the safety relevant parameters ‘crossing ghost island at the critical transition’, and ‘sum of (weighted) conflicts on the last 200 m of a passing section’. With regards to the determination of an optimal length of passing sections, the results of the analysis of platooning on feeder sections within 2+1 routes were used. It appeared that starting from a platoon share of 34.5% (according to the definition used within the frame of this investigation) platoons with at least seven vehicles in any platoon must be expected. It was shown that 15% of all platoons of seven vehicles required more than 900 m before they were dissolved. However, 95% of the 1,384 platoons where the dissolving process was observed until the end needed less than 900 m to dissolve. Furthermore, the frequency distribution of the 85th percentile platoon lengths revealed that considering peculiarities of individual feeder sections, a maximum 85th percentile platoon length of seven vehicles could be assumed.
As a result, both aspects together are proposed to avoid platoons longer than seven vehicles respective of the relevant share of platoons by restricting the length of the feeder section, and correspondingly the length of the passing section in the opposite direction. Since a correlation between the share of platoons, the traffic volume, the HGV share and the length of the feeder section could be proved, this specification serves to determine a maximum length of a passing section. Due to the observed effect of an increasing HGV rate leading to a decreasing maximum length (Figure 9), this implies that several shorter sections should be preferred to fewer longer sections for higher HGV rates. In this context, the term: ‘HGV-rate’ refers to the percentage of HGV traffic within the traffic stream. In principle, this coincides with the recommendations regarding an optimum length contained in the RAS-Q 96 (FGSV, 1996) derived from previous research results (Brannolte et. al., 1992).

6. RECOMMENDATIONS

6.1 General

Apart from recommendations for the determination of the optimal passing section length within a 2+1-route gained from this investigation, recommendations are provided for cross section design and choice of junction type which are partially based on findings from previous studies, but which have also been derived from available investigations in the frame of the dissertation (Irzik 2009) summarised in this paper.

6.2 Determination of the optimal length of a passing section

There are a number of marginal conditions that need to be observed in the application of the procedure developed under this study, because the investigated sections were specifically chosen because of their high level of comparability of the passing sections, regarding the determination of the required passing section lengths. As a result of the sections investigated, influences of the pre-set constant design and operational characteristics could not be examined.

The new approach for the determination of the optimal length of an passing section is predominantly only applicable to 2+1 routes which fulfill the fourth characteristic (slight gradient) described previously. As an additional marginal condition, it must be considered that passing was prohibited for HGVs on sections with a higher traffic volume, and simultaneously higher HGV rate. Traffic volumes between 400 vehicles per hour and direction and lengths below 800 m were excluded from the investigation and are also not within the application area for 2+1 routes as defined by the RAS-Q 96 (FGSV, 1996). A traffic volume of 1,300 vehicles per hour, per direction must be considered as upper limit. With a practical procedure for the determination of the optimal passing section length in mind, the specification of a quality standard for the number of un-dissolved platoons is suggested. Figure 9 shows the admissible recommended passing section lengths, depending on traffic volume, position within the 2+1 route (referred to as: ‘pos’ in the figure), HGV rate, and share of un-dissolved platoons.
In order to limit the lengths of passing sections with regards to platooning in the one-lane sections of the opposite direction, a platoon share of 34.5 % (according to the platoon definition) applied in this study is as advised as ‘to be avoided’ (see chapter 5). For such a platoon share, platoon lengths of more than seven vehicles must be expected. However, it was shown that platoons of such a length are rare, but often characterised by extremely long platoon dissolving distances when compared to the average. For this reason it seems to be admissible to limit the length of passing sections according to the chosen criteria in order to avoid large platoons.

A percentage of 32.5 % of vehicles, changing from the passing to the right lane at the end of a passing section, is seen to be critical and therefore must be regarded as ‘to be avoided’ in the determination of the minimal length of passing sections derived from safety considerations. This specification is guided by the fact that even though only for rates above 45 %, an increasing number of (weighted) conflicts per vehicle was observed, the range between 20 % and 45 % was not covered by the investigation. At a rate of lane changes at the end of 32.5 %, a (weighted) number of 0.036 conflicts within 30 minutes per vehicle must be theoretically expected. This limit is considered to be admissible because in approximately one-third of the 21 observed 30 minute intervals, less (weighted) conflicts per vehicle occurred within 30 minutes.

For sections beginning with an uncritical transition, a reduction of about 75 m of the required section length for complete platoon dissolving (determined according to figure 9), appears to be admissible due to the results obtained in the analysis of platoon leaders and the resulting differences with regards to the start of platoon dissolving (see chapter 5). In these sections also, the maximum length of a passing section can be 200 m longer because only the ghost islands located at the transition areas must be added to the length of a passing section to reach the length of the one-lane section of the opposite direction. For sections that begin with an added lane at a partially grade-separated junction, the junction area has to be considered in contrast to the comparatively short ghost island at an uncritical transition. This area was calculated to be approximately 200 m.

Figure 9 shows as an example the determination of the required length ($L_{\text{req}}$) for a passing section under consideration of the maximum length ($L_{\text{max}}$) and the recommended minimum length ($L_{\text{min}}$) for passing sections at position 1 (Pos. 1, see Figure 1) with a maximum traffic volume of 700 vehicles per hour and direction and a share of HGV of 10 %.
6.3 Cross section design

Within their study, Palm & Schmidt (1999) evaluated various lane and carriageway widths according to traffic safety by comparing various accident parameters. It became obvious that a lane width of 3.50 m, combined with a 0.50 m shoulder should be preferred for safety reasons. However, to enable a passing of road maintenance teams or stopped vehicles on single lane sections within 2+1-routes, a width of minimum 5.25 m should be provided. This would leave a 15 cm safety margin according to the legal vehicle width of 2.55 m as per the StVZO §32. The new German Guidelines for the Design of Rural Roads (RAL) (FGSV, 2008), currently being developed, therefore suggest installation of a shoulder of 0.75 m adjacent to the 3.50 m lane on one side, and a 1.00 m wide median on the other side (Figure 10). The wider median can additionally serve as a distinguishing criterion between the different design classes according to RAL, under the ‘self explaining roads’ philosophy. Guidance for the optimal design of this median will be provided via research presently being undertaken.

The investigation by Weber & Löhre (2003) showed that the widths of the verges in the single lane sections included in their investigation were below the 2.50 m minimum stipulated in the guidelines, however this was not reflected in the accident rates. Thus, a recommendation of the investigation was that it should be possible to narrow the verge width to only 1.50 m, without negative impacts on safety. However, it should be noted that if the section is located on a dam where there are crash barriers at the roadside, the verge width should be at least 1.80 m to enable the installation of steel crash barriers with the impact range W4 (=1.30 m) (1.80 m = 1.30 m + 0.50 m distance from road edge to crash barrier).
6.4 Junctions

Among the routes investigated by Weber & Löhe (2003), those with a relatively straight alignment and partially grade-separated junctions displayed the most favourable accident-cost rates and an even speed level. Thus, a high design standard should be intended for 2+1-routes wherever possible. It can be deduced from this finding that, if possible, uncritical transitions should be positioned in such a way that there will be an added lane in junction areas for the entering traffic. There were no substantial disruptions of traffic flow and traffic safety observed by Weber & Löhe for this junction layout.

Work performed by Kölle (1999) supports the construction of partially grade-separated junctions along 2+1 routes recommended by Weber & Löhe. Kölle recommends the implementation of partially grade-separated junctions wherever high travel speeds should be achieved for simultaneous short junctions distances and high junction volumes. Since 2+1 routes in general belong to highest level of road category for rural roads LS I (FGSV, 2008), high travel speeds should be guaranteed. Simultaneously, high junction volumes prevail on 2+1 routes.

The raised concerns that those ‘high’ speeds will be retained from the links into the partially grade-separated junction cannot be supported by the analyses of speed behaviour which were carried out within the frame of this investigation. Firstly, it must be stated that traffic reaches the partially grade-separated junction only following a one-lane section, not following a passing section. For one-lane sections (inside and outside 2+1 routes) the results showed that excessive speeds were observed in only a minor number of cases, and the legal speed limit was only typically violated in similar number of instances to that observed on ‘ordinary’ two-lane rural roads.
Furthermore the results derived from the investigation by Kölle (see chapter 3) showed that partially grade separated junctions on 2+1 routes hold a much higher safety level compared to the safety level of other junction types on 2+1 routes, but also compared to other partially grade-separated junctions on two-lane rural roads.

7. CONCLUSION

As a major contribution to the advancement of knowledge a practical procedure for the determination of the optimal length of a passing section on 2+1 routes was developed within the frame of the dissertation (Irzik 2009) summarised in this paper. In contrast to the procedure for the determination of a necessary length with regards to the dissolving of platoons as known from literature (Roos, 1989), it also contains recommendations concerning a minimum length derived from safety considerations as well as an upper limit in order to avoid excessive platooning in the opposite direction.

The newly developed procedure is not based on theoretical deliberations but on comprehensive empirical studies as well as correlation and regression analyses. Together with the procedure for the assessment of traffic quality on roads with 2+1 layout developed by Brannolte et al. (2004), and Baselau (2006), traffic planners have two helpful instruments for planning 2+1 routes.

While the procedure developed within the dissertation (Irzik 2009) that forms the basis of this paper serves to determine the section lengths in the pre-planning stage, the other procedures used by Brannolte et al. (2004), and Baselau (2006) can be used to assess the level of service on 2+1 routes.

In addition to the developed procedure the paper describes some new design standards in order to explain (cross section) and support (type of junction) a couple of suggestions in the new German Guidelines for Rural Roads (RAL).
REFERENCES
EVALUATING NETWORK WIDE TRAFFIC MANAGEMENT IN THE AMSTERDAM REGION

Status: Final
Date: May 19th, 2011

Suerd Polderdijk, MSc
Rijkswaterstaat Centre for Transport and Navigation
Mail: Schoemakerstraat 97c
2628 VK Delft
The Netherlands
Phone: +31 887982701
Fax: +31 887982999
Email: suerd.polderdijk@rws.nl
ABSTRACT

Congestion is a regular phenomenon in the Amsterdam area. In the project “Improving traffic flow on the A10” the Dutch national road authority and various local road authorities cooperate to reduce this congestion. Measures taken include the large-scale implementation of ramp metering on the on-ramps of the A10 ring road, various small infrastructural modifications, information panels on both urban roads and motorways and the integrated control of traffic management measures on both urban roads and motorways.

The effects of the project, on traffic flow as well as the opinions and behaviour of road users, are elaborately evaluated in two phases. In the first phase, the effects of ramp metering and small infrastructural modifications have been determined. In the second phase, the additional effects of automated coordinated control of traffic management measures will be determined. This paper describes the setup and results of the first phase of the evaluation of “Improving traffic flow on the A10”, which show that these measures have significantly improved traffic flow on the A10 motorway. Furthermore, this paper describes the setup of the second phase of the evaluation, which will determine the added benefit of integrated, network wide control of these measures.

1) **INTRODUCTION**

**A NETWORK-WIDE IMPLEMENTATION OF TRAFFIC MANAGEMENT**

Traffic management aims to optimize road network performance. This optimization is achieved by synchronizing infrastructure demand and supply, without large extensions of existing road infrastructure. Examples of traffic management measures are ramp metering, Variable Message Signs, Dynamic Speed Limits, and truck lanes.

“Improving traffic flow on the A10” is a large-scale implementation of traffic management measures in the Amsterdam area. In this project, Rijkswaterstaat, the Dutch national road authority, the urban region of Amsterdam, the municipality of Amsterdam and the province of North-Holland cooperate to improve traffic flow on the A10 Ring Road and main urban roads of Amsterdam. Traffic management measures have been implemented on the networks controlled by the various cooperating road authorities. These measures include ramp metering on the on-ramps of the A10, various small infrastructural modifications and information panels on both urban roads and motorways.

**INTEGRATED CONTROL OF TRAFFIC MANAGEMENT MEASURES**

Unique about this project is that these traffic management measures are connected and operate in an integrated, network-wide control environment. Currently, traffic management is mainly deployed “stand alone”, in an isolated manner (Taale, 2008). This means that traffic management measures generally relieve a single bottleneck and individual measures do not cooperate to improve total network performance. Such local or isolated traffic management may result in a suboptimal network performance, since capacity in networks is not utilized optimally (Mahmassani et al, 2005; Taale, 2008). The traffic management measures of “Improving traffic flow on the A10” however operate in a single control environment used by both Rijkswaterstaat and the municipality of Amsterdam. Traffic management measures are coordinated using a “Scenario Coordination Module”, which selects and implements the appropriate set of measures for the prevailing traffic conditions.

---

1 In Dutch: “Verbeteren Doorstroming A10”
EXTENSIVE EVALUATION

Numerous studies have been done worldwide to determine local effects of traffic management measures, such as ramp metering. However, few evaluations have been done that determine network wide effects of large scale implementations of traffic management measures. An example is an experiment by Minnesota DoT (2002) which shows travel speeds reduced by 7 percent and travel times increased 22 percent after shutting down all ramp meters in the Minneapolis-St.Paul area.

The extent to which integrated control of all measures in a network, for example with the “Scenario Coordination Module” in Amsterdam, improves network performance, in relation to its costs, remains unclear (Ministerie van Verkeer en Waterstaat, 2008). The practical applications of integrated control that do exist, are limited to one type of measure, such as coordinated ramp-metering (Papamichail and Papageorgiou, 2008) or coordinated intersection control.

An extensive evaluation programme has therefore been developed for “Improving Traffic Flow on the A10”. This evaluation consists of two phases. In the (completed) first phase the effects on network performance of local, stand-alone control of the measures was determined. In the second phase, the added effects of control using the “scenario coordination module” will be determined to show the added benefit of integrated, network wide control using traffic management scenarios. In the evaluation both the effects on the motorway network and urban roads are determined. Evaluating traffic management measures in such a large network required a specific evaluation method. This method includes extensive data collection and analysis and surveys amongst road users to determine road user behaviour and their attitude towards the measures.

PAPER OUTLINE

This paper describes the setup and results of the first phase of the evaluation of “Improving traffic flow on the A10”. Furthermore, this paper describes the setup of the second phase of the evaluation, which will determine the added benefit of integrated, network wide control of these measures using traffic management scenarios.

In chapter 2) the Amsterdam network and traffic management measures of “Improving Traffic Flow on the A10” are elaborated on further and the setup of the evaluation of this project is introduced. Chapter 3) shows the results of the first evaluation phase. Chapter 4) covers future work in the second evaluation phase and in the “Field Operational Test Integrated Network Management”, a large-scale follow-up of the current project. The final chapter presents the main findings so far and illustrates the relevance of the future work.
2) **Evaluation Setup for “Improving Traffic Flow on the A10”**

**Improving Traffic Flow on the A10**

The “Network vision North-Holland” describes the way the network in this province should function. This network vision was agreed upon by all road authorities involved: Rijkswaterstaat, the Dutch national road authority, municipalities and the province of North-Holland. Part of this network vision is a prioritization of the various links in the network. The A10 Ring Road, see Figure 1, received the highest priority. The underlying theory is that this ring road should function as a large roundabout and should remain free of congestion for as long as possible. If traffic on the “roundabout” comes to a standstill so will all traffic on connecting arterials.

![Figure 1: Map of the Amsterdam network, consisting of a ring road (A10), outer ring road (A9), an urban ring road (S100) and four main corridors connecting the S100 to the A10](image)

The project “Improving Traffic Flow on the A10” is a first step towards the implementation of the Network Vision North-Holland. Traffic management measures have been implemented on both urban roads and motorways. These measures consist of:

- Ramp metering on all on-ramps of the A10 motorway (32 installations). Each ramp metering installation is connected to the nearest traffic signal group on the urban roads. Together these measures aim to optimize traffic flow on the A10, while guarding that queues do not become too long or block crossroads.
Various small infrastructural modifications. These include modifications of road markings in a weaving area (one location) and the enlargement of streaming space at crossroads (6 locations).

Information panels on both urban roads and motorways. These information panels inform road users about travel times over alternative routes.

All traffic measures, including traffic signals on the urban ring road and the four main urban corridors, are controlled by a “scenario coordination module”. This module determines the prevailing traffic conditions on both urban roads and motorways and selects the appropriate set of traffic management measures, based on the priorities determined by the cooperating road authorities.

**EVALUATION SETUP**

This innovative project is extensively evaluated. The goal of this evaluation is to determine the impact of the project on the accessibility of the Amsterdam area and to verify whether hindrance on urban roads (such as increased rat running and blockings of crossroads) has not increased. Thus, effects on both motorways and urban roads are determined. Also, road user behaviour and their attitude towards the measures are researched.

The project is evaluated in two phases. The first phase of evaluation concerns the impact of the small infrastructural modifications and the ramp metering installations, connected to traffic signals. This evaluation was completed in the end of 2010 and chapter 3) will present the results. The second phase concerns the added benefits of the provision of route advice using information panels and the coordinated control of traffic management measures in the Amsterdam area using the Scenario Coordination Module. Results of this second phase are expected in the end of 2011. Chapter 4) will elaborate on this second phase. Using this phased approach, the additional impacts of coordinated control using scenarios compared to stand-alone control of traffic management measures can be determined.

**Measuring Impacts on Accessibility**

The “attribution question”. A major challenge in the evaluation is to determine whether measured changes in traffic flows can be attributed to the traffic management measures taken in “Improving traffic flow on the A10”. Why this is a challenge, is explained below.
“Improving traffic flow on the A10” aims to influence the accessibility of the Amsterdam area. It does this using traffic management measures, such as ramp meters and information panels. These measures send signals to road users (a red or green signal on a ramp metering installation or travel advice displayed on an information panel). These signals influence road user behaviour: they decide to stop for a red signal or change their route. This expected road user behaviour leads to various local effects, for example reduced inflow towards the A10 motorway from on-ramps equipped ramp metering. We assume that these local effects help to improve traffic flow on arterials and will lead to an increased accessibility of the Amsterdam area. The top half of Figure 2 illustrates this.

However, at the same time various other elements influence road user behaviour, the capacity of links in the network and traffic demand in the network and thus also influence accessibility of the Amsterdam area. The bottom half of Figure 2 mentions several of these elements. In the evaluation only the impact of the project is of interest, so these elements are considered “external influences”.

![Figure 2 Both the project and “external influences” affect accessibility of the Amsterdam area](image)

**Dealing with the “attribution question”**. The following approach was used to determine whether measured changes in traffic flows could be attributed to the measures taken (and not to some external influence):

- The expected effects of the project on local, arterial and network level were translated into hypotheses. For more on such an approach, see Polderdijk et al (2010). The idea is that effects on smaller network scales are easier to identify. If expected effects appear on smaller network scales, it is more credible that changes on a network wide scale can, at least to some extent, be attributed to the measures taken.
- The hypotheses were accepted or rejected using statistic tests to determine significant differences between the measurement periods. The datasets for these tests were selected in such a way that total inflow in the (sub)network studied was equal in both measurement periods. This excluded that increases or decreases in traffic demand influenced traffic flows.
- Data of periods and locations where external influences had a major impact on network capacity were marked and excluded from the datasets used for testing the hypotheses.
- Finally, results of the measurements were compared to the results of the road user evaluation (see below) and a simulation study (see Taale (2009)). This simulation study was done to predict in advance the impacts of the measures on traffic flows in the Amsterdam network.

**Data collection.** Various sources of data were needed for the approach described above. Data on motorways was collected using loop detectors. Data on urban roads was collected using loop detectors, cameras which measure travel times on Amsterdam’s main corridors (see Figure 1) and visual observation. Also, log files of traffic management measures were used. Data on external influences was also gathered, such as dates and locations of major road works, incidents and events. To limit the amount of data to be gathered and analyzed, the focus of the research was on the morning and afternoon peak periods.

The data described above is gathered during three measurement periods. The measurement period at the end of 2008, before implementation of the measures, was used as a reference period. A second measurement period was at the end of 2009 and beginning of 2010, after ramp metering and small infrastructural modifications were implemented. Currently, the route information panels and Scenario Coordination Module are operational and a third measurement period is between April and June 2011.

**Determining Road User Behaviour and Opinion**

Road user behaviour and their attitude towards the measures were determined using extensive internet surveys amongst road users. One thousand road users formed an internet panel, which was interviewed during each measurement period using an internet questionnaire. The road users had to be familiar with the A10 ring in order to be selected for the questionnaire. Road users were given questions about their travel behaviour, route choices, their opinion about the measures taken and their expectation about the effects of the measures.

Since the various measurement periods were several years apart (2008-2011), part of the road users left the panel during the study. They were replaced by new road users who met the criteria. It was checked whether there were significant differences in the answers of the “new” and original participants of the research.
3) **RESULTS OF FIRST EVALUATION PHASE**

**Impacts on Accessibility**

*Impacts on delay.* On an average working day the total amount of delay on the A10 ring road (including its on-ramps) decreased by over 10% due to the measures. For this calculation only days with a comparable total amount of inflow into the network were used. Data of periods when external influences had a major impact on network capacity were excluded. The increase of delay due to waiting times at ramp meters could not be directly measured and was calculated. This was done using on ramp meter log files, loop detector measurements and several (conservative) assumptions. The results are summarized in Table 1. The results are comparable to the simulation study of the measures, which reported a 10% reduction of total delay.

*Table 1: Traffic flow on A10 before and after implementation of measures*

<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
<th>After implementation of measures</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total vehicle kilometres on A10 [x1000 km]</td>
<td>3245</td>
<td>3243</td>
<td>- 0.1 %</td>
</tr>
<tr>
<td>Total delay on A10 [hours]</td>
<td>13388</td>
<td>11714</td>
<td>-12.5 %</td>
</tr>
<tr>
<td>Increase of delay caused by waiting times for ramp meters</td>
<td></td>
<td>+ 1.7 %</td>
<td></td>
</tr>
<tr>
<td>Conclusion: reduction of total delay due to measures taken</td>
<td></td>
<td>Over 10 %</td>
<td></td>
</tr>
</tbody>
</table>

*Impacts on travel times.* Average travel speeds increased on the A10 in the evening peak period. The increase of travel speeds varies per segment of the ring road from +3% up to +10%. Figure 3 illustrates where the largest improvements were found. On one segment a small decrease of travel speed was measured. No significant changes were found in the morning peak period. If one would drive a (fictitious) circle over the A10 ring road, one would experience -12.5% travel time in the clockwise direction and -7.0% in the counter clockwise direction due to the measures. On many routes using the A10, travel time reliability has increased.

*Local effects.* As indicated above, the effect of the measures varies per segment of the ring road. Some ramp meters had the expected effects and some ramp meters had no significant positive effect. This was due to the connected traffic signal on urban roads reporting potential blockings due to queues (which stops the ramp metering) or wrong parameter settings / malfunctioning of the ramp meter. Local effects were largest where a combination of ramp metering and small infrastructural modifications were implemented.
Urban roads. No significant increase of hindrance on urban roads was detected in the evaluation, based on measurements and observations of local traffic managers. However, less data was available for this aspect and it proved difficult to relate data to the measures taken. This conclusion is therefore more qualitative in nature. In the second phase of the evaluation, see chapter 4), an attempt will be made to draw more quantitative conclusions on this aspect.

Road User Behaviour and Opinions

The findings of the road user surveys supported the findings of the data analysis. Road users experienced an improved traffic flow in evening peak periods and a smaller improvement in the morning peak period. They also reported an improved traffic flow outside the peak periods. Road users did not report an increase of travel times on urban roads. To what extent these experienced improvements can be attributed to the measures and to what extent it is caused by other factors, such as a reduction of traffic demand due to the economic crisis, could not be determined.

The majority (63%) of road users also indicated that they support the concept of ramp metering and the majority (63%) also experienced that ramp metering improves on traffic flow on the A10.

Summary of Results – First Evaluation Phase

The results of the first evaluation phase demonstrate that the measures evaluated in this phase (ramp metering and small infrastructural modifications) improved traffic flow on the A10 ring road. On the whole, no increase of hindrance on urban roads was found. The improvements are mainly visible in the afternoon peak period, in the morning peak period no significant improvement was identified.
4) **Follow-Up: Second Phase and “Field Operational Test Integrated Network Management”**

**Second Evaluation Phase**

Between April and June 2011, data is gathered for the second phase of evaluation of “Improving traffic flow on the A10”. Data of this third measurement period will be compared to the second measurement period. Thus, a comparison will be made between stand-alone control of the measures and the coordinated control using the Scenario Coordination Module. This comparison will provide useful insights into the added benefits of coordinated, network wide control of measures using automated traffic management scenarios.

Another aspect which will be studied in the second evaluation phase is the added benefits of provision of travel information to road users. Route information panels have been installed around the A10 and in the city centre of Amsterdam. The traffic management scenarios controlled by the Scenario Coordination Module include route advice which is displayed on these information panels. The impact of such information provision will be studied using traffic data analysis and a survey amongst road users. Road users will be asked about their experiences and opinions of the information provided.

Results of the second evaluation phase are expected at the end of 2011.

**Field Operational Test Integrated Network Management**

Another follow-up in the Amsterdam area is the “Field Operational Test Integrated Network Management”. In this project, various types of coordination of traffic management measures will be developed, implemented and evaluated (Landman et al, 2010, Hoogendoorn et al 2010). This project’s goal is to further improve accessibility of the Amsterdam area and provide further insights into the added benefits of Integrated Network Management. Cost Benefit Analyses are part of the evaluation programme for this project.

A “Proof of Concept” for this project was completed at the end of 2009 (Landman et al, 2010).
CONCLUSIONS

Network wide implementations of traffic management require a specific and structured evaluation approach supported by structured data collection. Such evaluations yield useful insights into the impact of traffic management on a network wide scale.

The results of the first evaluation phase demonstrate that a network wide implementation of traffic management measures, in this case ramp metering at all on-ramps of the A10 ring, coupled with traffic signals and small infrastructural modifications, can have significant positive effects on network performance. A positive impact on the accessibility of the Amsterdam area was found: total delay on the A10 on average working days has been reduced by over 10% due to the measures of “Improving Traffic Flow on the A10”.

More insight is needed into the added benefits of integrated, network wide control of traffic management measures. The second phase of evaluation will provide valuable insights by determining the added benefits of coordinated control of the measures using traffic management scenarios and route advice provision to road users. The follow-up to the current project, the “Field Test Integrated Network Management Amsterdam”, will provide more insight into effects, costs and benefits of various forms of coordination of traffic management measures.

Acknowledgements

The impact evaluation was performed by DHV consultants, the road user evaluation was performed by I&O research and part of the data collection was done by Dufec consultants. The author would like to acknowledge Feiko van der Veen, Suzanne van Lieshout, Gerben Huijgen, Bart Dolstra and their respective teams for their work. Furthermore, the author gratefully acknowledges the contributions to the evaluation of Rob van Hout and Henk Taale and all road authorities involved.
REFERENCES


PERCEPTION, CHANGE DETECTION AND UTILISATION OF ROADS USING DYNAMIC INFORMATION: A THEORETICAL REVIEW

Drs. Ilse M. Harms (Corresponding author)
Rijkswaterstaat Centre for Transport and Navigation
Mail: Schoemakerstraat 97
2628 VK Delft
The Netherlands
Phone: +31 887982702
Fax: +31 887982999
Email: ilse.harms@rws.nl
ABSTRACT

Dynamic speed limits are part of the latest traffic management pilot program in the Netherlands in order to adjust real-time driver speed to the circumstances of the road condition. The general objective of this literature review is to understand the effectiveness of dynamic speed limits in terms of perception, as little is yet known how people perceive changes in dynamic speed limits in relation to prevailing road conditions. Dynamic speed limits are displayed on information carriers with the intention to change the drivers’ speed. A number of previous studies on change detection, such as Rensink (2002a), have shown that ‘under a wide variety of conditions we can be amazingly blind to changes, failing to see them even when they are large, repeatedly made and anticipated’. This phenomenon has been reverted to in literature as change blindness (Rensink, 2002a). Dynamic speed limits displayed on information carriers compete with a wide variety of other types of information as previous studies have shown the human information-processing system is limited in its ability to deal with the amount of information displayed (Wickens et al., 2004). Dynamic speed limits are considered road information. They are intended to inform drivers about the status of the road and the goal is that drivers should adapt to them in real-time. In this paper a framework is suggested to help researchers distinguish between route information, road information, road user information and driving unrelated information.

This literature review shows that information especially prone to change blindness, is information that lacks perceptible transient motion signals when changing (e.g. due to masking or gradual change), is not relevant for the task at hand, or violates expectations about objects of significance for performing the task at hand. Both transient motion signals attracting attention, as well as the task at hand, influence the visual search needed to build an internal representation of the environment. Based on the current literature review, it can be concluded that detection of changes in dynamic speed limits will suffer from gap-contingency masking and changes in semantic identity. It remains unclear whether repetition of successive information carriers will reduce the chance of detection of changes in dynamic speed limits. Detection of these changes might be increased as they are necessary to maintain an appropriate speed; this is part of the driving task.

In conclusion, some drivers will not detect changes in dynamic speed limits, but it remains unclear to what extent this will be the case. Subsequent research should establish the extent of the change blindness effect and whether it can be reduced to increase situation awareness and the effects of dynamic traffic management.

Keywords: change detection, change blindness, visual search, perception, situation awareness, dynamic speed limits, variable speed limits, VSL, traffic management.
1. INTRODUCTION
Since the late seventies the Netherlands have used traffic management as a tool to influence traffic behaviour and traffic flows. In 2009 the Dutch road authority has started experiments with dynamic speed limits on four designated motorways, to real-time adjust driver speed to the circumstances. This is one of the most recent pilot programs in the Netherlands, which has the potential to be implemented nationwide as part of the current Dutch traffic management philosophy to utilise road infrastructure to ensure ‘the best possible handling of traffic demand’ (Ministry of Transport, 2008). Utilisation means that the road authority tries to inform, advise, guide and steer road users to influence their autonomous behaviour. To do so, information must be real-time, accurate and dynamic, what means that it will change over time.

1.1 What do we know about change detection of dynamic speed limits?
Research on the Dutch pilot study with dynamic speed limits (called Dynamax), showed that changes in dynamic speed limits altered driver speed across a longer period of time (Burgmeijer et al., 2010). When this time frame is reduced to the actual moment of speed limit change, on average drivers do not immediately increase their speed when allowed (Harms & Brookhuis, 2010). This suggests that change detection might be an issue. Hoogendoorn (2010) showed that the position of the speed limits in the pilot program Harms & Brookhuis (2010) used, also affects detection. In the pilot, dynamic speed limits were presented on high poles on each side of the road, instead of on gantries above the road. Hoogendoorn (2010) showed that information on high poles is less likely to be seen, compared to information positioned over the road. These findings are supported by Chang & Li (2008), who also found overhead speed limits to be more easily detected, compared to other configurations of dynamic speed limits.

Although the position of dynamic speed limits can help optimise detection, it does not answer the questions whether and when drivers detect changes in dynamic speed limits. While research has been done on different configurations of dynamic speed limits, no research could be found on the detection of changes in dynamic speed limits (also called variable speed limits or VSL). These speed limits are highly relevant for the driving task, visible when drivers need to respond, but they also change over time. Previous studies on change detection showed that ‘under a wide variety of conditions we can be amazingly blind to changes, failing to see them even when they are large, repeatedly made and anticipated’ (Rensink, 2002a).
1.2 Will changes in dynamic speed limits be detected?

The general objective of this literature review is to understand the effectiveness of dynamic speed limits in terms of perception: how likely is it that changes in dynamic speed limits suffer from change blindness? To tackle this subject of visual change detection, it has been divided in two research questions; do drivers perceive driving-related information on information carriers, and do they perceive changes in driving-related information on information carriers?

2. PERCEPTION OF DRIVING-RELATED INFORMATION ON INFORMATION CARRIERS

To answer the question of perception of driving-related information on information carriers, it needs to be understood how perception works when people drive. Perception ‘involves the extraction of meaning from an array (visual) or sequence (auditory) of information processed by the senses’ (Wickens et al., 2004). Like dynamic speed limits, the information drivers get is still largely visual. Although it has been estimated that 90% of the amount of information drivers use is visual (e.g. Hills, 1980), there is no exact percentage for it (Sivak, 1996). However it is widely agreed that relevant driving information is predominantly visual. Vision is therefore considered the most important sense for the driving task (Wickens et al., 2004; Hills, 1980; Sivak, 1996). Since driving is a highly visual task, it depends predominantly on the visual information that drivers look at and actually see. Detection and apprehension of visual information plays a major role in maintaining situation awareness. In order to evoke a response, visual information must be selected and processed.

2.1 What information do drivers come across?

Drivers get a lot of information while driving on a road, outside as well as inside the vehicle. This information may be visual, auditory or tactile on topics which vary from today’s weather forecast to road closures. As the human information-processing system is limited (Wickens et al., 2004), information on information carriers has to compete with other information, which might not even be necessary for the driving task or even related to this task. Driver information can be categorized in different types of information, based on Michon’s task hierarchy of driving (1985). He divides driving in tasks at a strategic level of control (e.g. route choice), a tactical level of control (e.g. braking for a crossing pedestrian) and an operational level of control (e.g. shifting gear). When this model is applied to driver information, a differentiation can be made between route information, road information, road user information and a separate category for driving unrelated information.
- **Route information.** Information that tells drivers in advance how to get from A to B and what they might come across along the way, like partial road closures or traffic jams. This information is needed for general trip planning, a task performed at Michon’s strategic level of control. The information applies to tasks that may take several to many minutes and includes route guidance by signposting, as well as ‘orders’ from a navigation system. It enables drivers to change their route, destination or modality well beforehand if necessary.

- **Road information.** Information which informs drivers about the current status and state of the road, information necessary mostly for Michon’s tactical level of control, but the operational level too. The status tells you which rules are in effect, like current speed limits and overtaking prohibitions, and warns you in advance for possible dangers like the tail of a traffic jam or a sharp curve. Information about the state of the road involves the current state of the road surface and its surroundings. Road information can come from the road itself (e.g., delineation or cracks in the road surface), road signs and in-car devices and applies to tasks that only take seconds to perform.

- **Road user information.** This concerns feedback about your own driving performance itself and in relation to other drivers. The main use of this information is on the operational level of control, though some might also be used at the other control levels. Generally, actions upon this type of information require just milliseconds. Road user information gives information about your own driving skills, for example, through a rumble strip or a lane departure warning assistant (LDWA) when line-crossing, and your driving capacity. Examples of the latter may serve as warnings for fatigue or mental workload overload. Examples of devices that give road user information are adaptive cruise control, an overtaking assistant, speed alerts concerning your own speed, a pedestrian alert and warnings for cars which travel the wrong way or warnings for broken down vehicles ahead. Most of this type of information comes from in-car devices.

- **Driving unrelated information.** This covers a wide range of information which is not necessary for the driving task at hand. Examples are phone calls about the business deal you are about to close, talk shows on the radio, screaming children in the back seat and general advertisements along the road.

Driver behaviour itself can be influenced by all the four types of information. To handle traffic demand as well as possible, traffic management information is used to influence driver behaviour through both route and road information, which may be displayed on information carriers along the road. Dynamic speed limits are considered road information; they inform drivers about the status of the road and to adapt to them should only take seconds. As dynamic speed limits are visual, they do not only compete with other types of information, they also compete for detection and processing through one channel (Wickens et al., 2004).
The eye is however not a camera. Scenes around us do not result in a picture-like representation in our brain, people do not inspect and fixate every part of a scene. Instead, we build internal representations of the scene around us (Rayner, 2009; Rensink, 2000). Perceptual errors can occur during the process of visual search and during the processing of selected information. These errors may even be worse when information on information carriers changes, which can result in change blindness.

### 2.2 How does visual information get selected?

Visual search precedes selection and consists of saccadic movements of the eye and fixations. It is only during these fixations that the eye holds still and acquires information visible at the point of fixation. During the time the eye holds still, called fixation time, the selected information is processed. Fixation times while driving are generally very short and vary from 250 to 600 ms. Information that is relevant, complex or informative may cause a longer fixation time and/or may cause the eye to fixate the information more often, and while the eye moves again it cannot extract information from the environment (Theeuwes, 2008). When does information get selected? Where do we, and drivers in particular, look? Visual search may be either invoked voluntarily by attention guided to a certain location e.g. based on experience and expectations (top-down selection) or attention drawn involuntarily to a salient stimulus (bottom-up selection) (Wickens et al., 2004; Theeuwes, 2008). Mostly visual search is a combination of both top-down and bottom-up factors, though a purely bottom-up, salience-based search may occur (Lamy & Zoaris, 2009).

### 2.3 Salient objects may still ‘catch the eye’

**Bottom-up selection.** When a scene is inspected without searching for a predefined object the most salient objects will attract eye fixations (Underwood et al., 2006). An object can be salient because it stands out relative to surrounding objects (Itti & Koch, 2001), for example due to flashes, movement, unique shape, colour or luminance (Wickens et al., 2004). Salient objects in traffic can be a blinking variable message sign, brake-lights or a single yellow fluorescent work-in-progress sign, or a bright billboard (Edquist et al., 2010). The most salient objects in a scene are not necessarily the first locations that will be inspected when freely scanning, although they are likely to be inspected within the first few fixations (Underwood et al., 2006). Bottom-up selection depends predominantly on features of objects in the scene. Objects that are salient will capture attention and ‘catch the eye’ to fixate them. However, several experiments showed that a well-specified task changes visual search patterns of scenes (Underwood et al., 2006; Pearson & Schaefer, 2005; DeAngelus & Pelz, 2009; Yarbus, 1967). Underwood et al. (2006) found that participants in search of a predefined target are less influenced by salient distracter objects. Their
participants fixated considerably less on the high saliency distracter when they had the task to search for a low saliency target, than without this task. The distracter object was still fixated in 20.4% of trials, compared to 84.5% of trials with the specified search task. Hence, cognitive override did not completely suppress attentional capture. DeAngelus & Pelz (2009) also found that overall their participants viewing patterns changed with instruction, although shifts in these viewing patterns were only very small for some participants. This was especially true for the participants with shorter viewing times. This brings up the question to what extent it is possible to cognitively override attentional capture and whether everyone is able to do so. The answer to this question is still under debate, for reviews see Lamy & Zoaris (2009). Lamy & Zoaris (2009) conclude that stimulus salience still guides attention irrespective of the task at hand. Despite the amount of power top-down control can have in the suppression of salient distracters, when freely viewing a scene salient distracters can result in failing to look at other objects in the scene.

2.4 Top-down controlled search may prevent information to be seen

Top-down selection. Goal-directed, or top-down visual search highly depends on expectations. They guide the eye to the locations where specific information is expected. These expectations can be distinguished in task dependant expectations and object dependant expectations (Martens, 2007). An example of the effect of task dependant expectations is given by Richard et al. (2002), who showed participants who watched images of driving scenes are slower to detect driving-unrelated changes than driving-related changes. Object dependant searches, are searches in which people have expectations about where to find the object and hence look at the likely locations. For example, when looking at a picture of a living room in search of a banana, it is more likely to be found in the fruit basket on the table than hanging from the ceiling.

The expectations a driver has are based upon earlier experiences. To organise past experiences and/or past reactions, we use schemata. Schemata, first introduced by Bartlett (1932), are active knowledge structures about particular topics in long-term memory. When drivers gain more driving experience, they obtain more elaborate schemata on what to expect when driving and where to look for specific information. The stronger the driving-related schema, the stronger the top-down controlled search for specific driving-related stimuli (Borowsky et al., 2008). Visual scanning patterns of experienced drivers differ from these of novice drivers, with experienced drivers showing greater overall sensitivity (Underwood et al., 2003). These well-learned schemata become automated. In general it takes hundreds of trials of training of consistently mapped relations to obtain full automaticity (Schneider & Chein, 2003; Shiffrin & Schneider, 1977). This may lead to top-down controlled automatic detection. Though, as Ranney (1994) points out, the question remains to what extent automatic detection of conspicuous stimuli in
driving results from top-down control through extensive learning (Schneider & Shiffrin, 1977), or from bottom-up control through attentional capture induced by distinctive stimuli features (Treisman & Gelade, 1980).

The well-learned schemata experienced drivers have, may also become a barrier when information from the driving environment conflicts with the schemata. Several studies showed errors in visual search when traffic signs violate expectations because they are located at an unanticipated location. These traffic signs were less likely to be detected and if detected, search took considerably more time (Borowsky et al., 2008; Hoogendoorn, 2010; Theeuwes & Hagenzieker, 1993), with the first to result in erroneous situation awareness. To unlearn this automatic-attention response is very hard, as it takes a considerable amount of retraining to learn new sets of automatic responses (Shiffrin & Schneider, 1977).

2.5 Strong expectations may prevent information to be processed leading to erroneous situation awareness

Successful selection of information is succeeded by processing. Well-learned schemata not only have a role in selection of information but also in it’s processing. The schemata will fill in blanks when only a few stimuli have been selected or when the selected information from the environment is incomplete. With the schemata the selected information is enriched into a complete picture (Proctor & Dutta, 1995). An example is reading only a few words from a sentence and still being able to grasp what the sentence is about. So strong expectations do not only affect where drivers look, they also affect processing of what drivers look at. When a driver looks but fails to see (s)he actively searches a visual scene, fixates on a stimulus but does not process it (Staughton & Storie, 1977; Galpin et al., 2009; Martens, 2007). It appears that information changed, is especially vulnerable to the error of looking but failing to see. This can occur when drivers are very familiar with a certain road and the information presented is unexpected. Martens & Fox (2007) showed that drivers who where trained to become highly familiar with a specific road failed to timely recognise that priority at a certain intersection had changed. Only two of twelve participants realised something had changed when they were already crossing the intersection. The other participants never noticed. Martens & Fox (2007) established that this error occurred due to the failure to look. Participants did look at the road sign showing the changed priority, but had such strong expectations about it because of their familiarization that they failed to apprehend. ‘Seeing’ information that is not actually there also leads to an erroneous situation awareness.

This erroneous situation awareness caused by strong expectations can in some cases also result in fatal accidents. An example is an accident with a tram colliding with a passenger car on December 11 in 1988, resulting in two fatalities. Both drivers stated their traffic light gave them
right of way, while research showed the traffic control system did not activate conflicting traffic flows simultaneously (Muller & Verweij, 1991). Although a traffic light provides dynamic information, drivers can have strong expectations about the content of this information. In this particular case, extensive prior experience with the traffic control system could have led to the strong expectation that the tram always had priority in time when approaching the intersection. Muller & Verweij (1991) showed that only in a small percentage of occasions this expectation would be violated. This might have resulted in the tram driver looking at the traffic light, but failing to see it did not give him priority.

2.6 Conclusions for perceiving information on information carriers

In summary, it is established that information on information carriers competes with other (types of) information. To evoke a response, this information must be selected as well as processed. When visually searching their surroundings, drivers select information while driving, a process that is mostly a combination of bottom-up and top-down driven search. When visual search is inaccurate, drivers may fail to look and information will not be selected. Failing to look can occur when the eye was caught by other, more salient information, or when the location of information violates expectancies about its location. Strong expectations might also prevent processing of information that violates these expectations, resulting in drivers ‘looking but failing to see’. Both perceptual errors may prevent perception of information on information carriers. ‘Failing to look’ as ‘looking but failing to see’, may lead to erroneous situation awareness.

To be perceived information on information carriers should be visible. To overcome the problem of failing to look, visual attention can be attracted either bottom-up or top-down. Visual attention can be actively drawn, forcing bottom-up selection, by making the information carrier more salient, e.g. by using flashers. This way of attracting attention should however remain exclusive for warning or steering drivers in order to maintain conspicuity, causing information that aims to inform, advise or guide drivers relying solely on top-down visual selection. Road authorities could also opt to make surrounding information less salient. Placing information where it is expected though, can mostly solve the problem of drivers failing to look. Overhead information is likely to be seen. To prevent drivers from looking but failing to see the information on information carriers, the information should match with drivers’ expectancies about it.

If located and designed well, driving-related information on information carriers is more likely to be perceived. But will this information still be detected if it changes occasionally?
3. CHANGE DETECTION OF DRIVING-RELATED INFORMATION ON INFORMATION CARRIERS

Failing to detect a change in a visual environment can be due to ‘failing to look’, by improper visual search, or ‘looking but failing to see’, by improper processing. Change detection is referred to as ‘the visual process involved in first noticing a change’ and consists of the detection of the change (‘i.e., the observer reporting on the existence of the change’), as well as the identification (what changed) and localization (where is it) of the change (Rensink, 2002a).

3.1 What is the effect of changes on perception?

Although we might think we see everything, there are several conditions in which we are remarkably blind to changes around us. An example is given by a video of Wiseman (2007). In this video Wiseman performs a magic trick that will change the colour of a deck of cards while being assisted by his co-presenter. The viewer’s attention is mainly focussed on the cards, trying to unravel the trick. In the meantime the shirts of both presenters, the tablecloth and even the curtain in the back change colour as well. All changes are performed out of sight of the camera, with very few viewers detecting the changed fabrics. The changes can be considered unanticipated or not part of the task (unravelling the magic trick), but Rensink et al. (1997) showed that even changes that are large, repeatedly made and anticipated can go unnoticed. Many studies of change detection have shown that participants are very insensitive to small as well as large physical differences in two succeeding images (e.g. Enns, 2009; Simons & Rensink, 2005; Rensink, 2002a; Simons & Levin, 1997; Rensink et al., 1997; O'Regan et al., 1996; Grimes, 1996). When the images are not interrupted and are shown successively, participants detect the change in the images immediately (Martens, 2007). Changes come with transient motion signals caused by the movement of the change. These motion signals can capture attention (Rensink, 2002a), attracting attention bottom-up. When transient motion signals are visible, attended and not disrupted, changes can easily be detected (Zheng & McConkie, 2010).

Change detection errors occur when transient motion signals can not be seen, for instance due to a visual disruption. This can either be a physical obstruction or a simultaneous presentation of distracters. Only a single change can be attended at any moment in time. This is also called ‘change simultagnosia’ (Rensink, 2002b). Particularly with driving, transient motion signals are easily masked. Change blindness can be characterized as ‘a failure to notice or become aware of a change in the environment. Typically due to a situation that masks the features of the objects in the environment that have changed’ (Hannon, 2009). It distinguishes from inattentential blindness in that the latter typically results from ‘focussed attention on a specific task’ (Hannon, 2009).
When drivers fail to fixate the changed stimulus again after the visual disruption, they make the error of failing to look. When failing to look, information is unattended which leads to an incomplete internal representation of the situation. In order to shift attention to this location visual search has to be top-down (Schankin & Wascher, 2008). When looking but failing to see, information is not sufficiently processed. This seems to be due to strong schemata repressing (the need to process) the actual information hence resulting in an incorrect internal representation. There is strong evidence that change detection errors occur due to the lack of richly encoded internal representations, which are necessary to do an adequate comparison between the ‘before’ and ‘after’ representation (see Hannon (2009) for reviews).

### 3.2 An overview of types of changes seen and not seen

So which changes do drivers see and which don’t they see? Various studies have been done in order to differentiate changes that are seen, from those that are not seen. Experimental conditions that can be distinguished are masking of objects, gradually changing objects, changing existence of objects (additions and deletions), changes within an object, changes in object location, object similarity and repetition, incongruence, movement of objects, center of interest, task-dependency and expectancies about objects.

**Masking.** Since masking was the main method used for studying change blindness, there is a large body of research on various types of masking all resulting in deteriorated detection of changes. Different forms of masking that can result in change blindness are:

- gap-contingency, when the change happens during a temporal gap (Rensink, 2002a);
- saccade-contingency, when the change happens during a saccade of the eye (Grimes, 1996);
- blink-contingency, when the change happens during an eye blink (O'Regan et al., 2000);
- splat-contingency, when the change happens simultaneous with the appearance of brief distracters, like ‘mud splashes’ (O'Regan et al., 1996; O'Regan et al., 1999) or the sudden onset of a lead vehicle’s brake lights;
- occlusion-contingency, when the changing item is briefly occluded (Rensink, 2002a), e.g. by a brief blank (Rensink et al., 1997; Galpin et al., 2009) or a passing a truck.

**Gradual change.** Not only objects that are masked but also objects that change gradually are susceptible to change blindness (Simons et al., 2000). Gradual changes are more difficult to detect than changes masked by a visual disruption. David et al. (2006) found that detection rates for gradual changes were three times lower compared to visual disruptions, with only 15% of participants detecting slow, gradual change in facial expressions.
'Existence' of an item (Rensink, 2002a); i.e. additions and deletions. Research suggests that some deletions are more easily detected than additions (see Rensink (2002a) for reviews), though both can suffer from deteriorated change detection. Detection of changes in existence appears to be one of the more noticeable types of changes, when motion transients are not available. Mondy & Coltheart (2000) compared changes in deletions, additions, colour and location, using photographs of naturalistic scenes varying in subject matter. They showed that identification accuracy was highest for deletions and lowest for changes in location. When motion transients were available, detection and identification of all these changes were equal. These findings agree with research of Pearson & Schaefer (2005), who used pictures of traffic related scenes. They also found deletions of objects to be detected more often than positional changes of objects. Although these findings give insight in detection of changes in existence of an item, it is important to notice that the results are based on changes in static images rather than dynamic images. Driving means travelling through a dynamic scene. Only few research has been done on changes in existence in traffic scenes, with participants actually driving through the scene. Martens & Fox (2007) showed that adding a yield line, also called Shark’s teeth, to a changed priority of an intersection resulted in ten out of twelve acquainted drivers not noticing the change and only one driver actually noticing the change, knowing what had changed but unable to react to the change timely. Charlton & Starkey (2010) reported mixed effects on both additions and deletions from a driving environment with which participants were strongly familiarized. Deleted road markings, which were foveal, were detected while the deletion of a prominent farm silo, which was located peripheral, remained undetected. Findings for additions were rather similar, with foveal stimuli like the addition of a police car being detected, while more peripheral stimuli went undetected, like the addition of a warning sign for a road dip. Hence the evidence for the detection of changes in existence in dynamic scenes is inconclusive. Also, in both studies participants are very familiarized with the driving environment prior to changes in existence occur. This familiarization is a confounding factor, making it difficult to explain the outcomes in terms of detectability of changes in existence in dynamic scenes. However, if change detection in dynamic scenes is similar to static scenes, the mere addition or deletion of an object in a scene appears not to be sufficient to induce a change detection problem.

*Change within object.* Changes within objects can be distinguished in changes in properties and changes in semantic identity (Rensink, 2002a). Property changes can be changing the size, shape or colour of an object, and detection can differ for different properties. Detection is best for changes to a unique property and worst for switches between objects (see Rensink (2002a) for reviews). When the meaning of a scene or object is not changed, changes to whole objects are identified more often than changes to objects which are part of a larger object (Mondy & Coltheart, 2000).
With changes in semantic identity the overall appearance of an object remains the same. Detection of this kind of changes is generally quite poor (see Rensink (2002a) for reviews). Recent examples in driving that support this assertion are the studies of Charlton & Starkey (2010) and Martens & Fox (2007). Charlton & Starkey (2010) found that changing the wording of a directional sign from English to German is not detected by participants who are very familiar with a specific route. Also changing a road sign with which participants were highly familiarized from priority to ‘give way’ resulted in none of the participants detecting the changed sign (Martens & Fox, 2007).

**Location.** Relocating an object in a scene appears to yield the worst change detection rate when the change lacks transient motion signals. This was found in studies comparing different types of changes, using naturalistic scenes (Mondy & Coltheart, 2000) as well as traffic scenes (Pearson & Schaefer, 2005). When the object changed was of marginal interest, the difference between identification of deletions and relocations disappeared as well (Pearson & Schaefer, 2005). Research of Simons & Chabris (1999) suggests that spatial proximity of an unattended target to attended locations does not influence detection. This is also supported by O’Regan et al. (2000), who found that fixation of the change location (within 1 degree) still led to participants failing to see the changes more than 40% of the time, for both central and marginal changes. This may indicate that observers attend to events and objects rather than spatial positions, as also pointed out by Simons & Chabris (1999). Based on the research mentioned above, changes in location might not to be noticed very well, as bottom-up visual search does not appear to be location oriented.

**Object similarity and repetition.** Objects which share similar features with target objects are more easily detected (Treisman & Gelade, 1980). Research of Simons & Chabris (1999) suggests that similarity to target objects increases the likelihood of detection of unexpected objects. Most et al. (2001) reported similar results, stressing the role of attentional goals of the observer in detection. They also add that objects which features differ greatly from ignored items have an increased likelihood of detection as well. Difficulties arise when objects that are completely similar to other objects, are changed. Research of Mondy & Coltheart (2000) suggests that repetition of an object results in deteriorated detection of changes in only one of these objects. Diminished change detection occurred for both deletions and additions which were part of a larger object as well as entire objects, when in the presence of an identical object.

**Incongruence.** Martens (2011) shows that changing the meaning of an object can affect change detection when it becomes incongruent with its surroundings. Martens changed an appropriate traffic sign in a driving scene into a traffic sign which meaning did not fit the scene. The latter
change was more often detected than changing the traffic sign into a sign that did match the scene. This finding of incongruency is not supported by Charlton & Starkey (2010), who changed the wording of a directional sign from English to German while driving on a New Zealand road. This suggests that congruency is not a factor in change detection on its own.

**Motion in space.** Silencing is ‘the motion-induced failure to detect change (Suchow & Alvarez, 2011). Suchow & Alvarez (2011) report effects of silencing on change detection, with motion suppressing the detection of changes in hue, luminance, size or shape of the target stimulus. The target they used consisted of a circle of dots positioned around a central fixation point. Their six participants knew the target was changing in hue, luminance, size or shape. When rotation speed of the circle of dots increased the detection of their feature changes decreased, with a rotation speed of 0.33 Hz resulting in almost complete silencing. Suchow & Alvarez (2011) argue that this motion induced failure to detect changes is not caused by motion in space but is induced by motion on the retina. When participants are able to follow the moving target with their eyes, change detection rates improve. Cole & Liversedge (2006) found that objects which were looming towards the viewer were detected more often than objects receding from the viewer. When observers are moving, instead of targets, Wallis & Bülthoff (2000) found that observer motion had a negative effect on the detection of changes in location and orientation of targets, which could not be found for changes in colour or existence. This suggests that mere changes in presentation do not solely induce change blindness effects that can be found in the real world, as interaction effects with motion do exist.

**Meaningfulness.** Alterations to items of central interest are more often detected than alterations to items of marginal interest (Pearson & Schaefer, 2005; O'Regan et al., 2000; Rensink et al., 1997). Pearson & Schaefer (2005) found this to be true for both changes in location and deletions. Using an eye-blink paradigm, O’Regan et al. (2000) found that changes to items of central interest are also detected earlier than changes to items of marginal interest. Changes to items of central interest were most often detected on the first blink. Whether an object is of central or marginal interest was tested before the change blindness experiment, by asking naïve observers to give a brief verbal report of the scenes to be used. Objects are considered of central interest when they are mentioned by more than half of the observers (three or more out of five observers in the experiment of Rensink et al. (1997) and O’Regan et al. (2000), and six or more out of nine observers for Pearson & Schaefer’s (2005) experiment). Rensink et al. (1997) and O’Regan et al. (2000) considered objects to be of marginal interest when they were listed by none of the observers, while Pearson & Schaefer’s (2005) where listed by at least one but fewer than four observers. The above mentioned findings concerning change detection of central interest items are supported by David et al. (2006). They found that changes in facial expression, which are of
particular importance to the visual system (see David et al. (2006) for reviews), are more often detected than comparable colour changes.

**Task-dependency.** As stated above, detection of changes in objects which suffer from visual disruption or which change gradually relies mainly on top-down controlled search. The change in these situations is just not conspicuous enough to capture attention. Top-down control can be induced by the task at hand. Emphasising the driving aspect in a one-shot photo change detection task resulted in participants performing better on detecting driving-relevant alterations than participants without this instruction (Pearson & Schaefer, 2005). Supporting this, participants who were watching pictures of traffic scenes were quicker to detect changes in targets semantically relevant for the driving task than in irrelevant targets (Galpin et al., 2009). Wallis & Bülthoff (2000) showed that the task at hand, in this case driving, also narrowed the main field of attention to the vicinity of the road. They used targets which were not relevant for the task, though changes in these targets were more likely to be detected by drivers when the targets were located on the road compared to near the road or even further away. These differences disappeared when participants were travelling through the same scene as a passive passenger rather than an active driver. The task at hand and instructional set can diminish change blindness if the targets are relevant for the task, but it can also increase detection for targets located near the main field of attention for the task. Task-dependant detection of objects is closely linked with expectancies, inducing top-down controlled visual search.

**Expectancies.** Changes which are not expected demonstrate stronger change blindness effects than expected changes (Simons & Rensink, 2005; Hannon, 2009). However, when strong expectations are violated, Martens & Fox’s study (2007) showed that the change is unlikely to be detected. Martens & Fox (2007) familiarized drivers with a specific route by repetitively driving the same route in a low-cost driving simulator. With increased exposure the recollection of traffic signs along the route increased while glance duration at the traffic sign decreased. At the last drive the experimenters changed priority at an intersection, adding a yield line and changing a priority traffic sign into a ‘give way’ traffic sign. The changed traffic sign did not provoke longer glance durations. The priority change was not detected in time by any of the participants. Only two out of twelve participants noticed a change, but both did not see the changed traffic sign. Just one of them saw the added yield line although he still wasn’t able to react timely to them. The other one noticed something had changed but did not know what it was. This is in line with Charlton & Starkey (2010) who also found drivers having problems with complying to noticed changed conditions, once driving in a specific environment became proceduralised due to extensive training and subsequent familiarization. Based upon their research, Martens & Fox (2007) argue that change blindness is stronger when very strong expectations about the stimulus
are combined with gap-contingency masking. This results in drivers looking, but failing to see the change.

### 3.3 Conclusions for perceiving changes in information on information carriers

In summary, it can be established that, like visual search, detection of changes is a mixture of bottom-up and top-down processes. Correct detection and identification of changes strongly depend on the task at hand and on expectations about objects of significance for performing this task. From the bottom-up perspective changes need transient motion signals or need to be large enough to capture attention. Small changes like changing the semantic identity of an object while the overall appearance remains the same are unlikely to be noticed, which can be amplified by strong expectations about the semantic identity. The task at hand highly interferes with the detection of objects as it changes visual search, therefore also affecting change detection. Table 1 gives a summary of the types of changes and their effects on change detection. In the third column of table 1 their relevance for information carriers is discussed.

<table>
<thead>
<tr>
<th>Type of change</th>
<th>Effects on change detection</th>
<th>Relevance for information carriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masking</td>
<td>The lack of a transient motion signal that results from masking deteriorates change detection.</td>
<td>Very relevant. As the driving scene is dynamic, all types of masking can occur. Gap-contingency masking is most relevant for information carriers an dynamic speed limits in particular, as drivers drive past information that is repeated on subsequent information carriers. This results in a temporal gap between information displayed.</td>
</tr>
<tr>
<td>Gradual change</td>
<td>The lack of a transient motion signal that results from gradual change deteriorates change detection even more than masking does.</td>
<td>Relevance depends on the type of information carrier used: changes in information are either gradual or abrupt. The latter is more prevalent in the Netherlands.</td>
</tr>
<tr>
<td>Existence of an item; i.e. additions and deletions</td>
<td>The mere addition or deletion of an object in a scene is not sufficient to create problems with change detection.</td>
<td>Not relevant, as just the information displayed changes, not the information carriers themselves.</td>
</tr>
<tr>
<td>Change within an object</td>
<td>Especially detection of changes in semantic identity is generally quite poor. Detection is best for changes to a unique property (e.g. size, shape, colour).</td>
<td>Very relevant. Changes in information on information carriers are mostly changes in semantic identity.</td>
</tr>
<tr>
<td>Location</td>
<td>Detection of changes in location is poor. Spatial proximity to attended stimuli does not increase change detection.</td>
<td>Not relevant. The location of information carriers is mostly static.</td>
</tr>
<tr>
<td>Object similarity and</td>
<td>Target similarity increases detection.</td>
<td>Relevant for road information, which is prone to</td>
</tr>
</tbody>
</table>

Table 1. Types of changes: a summary of effects on change detection and relevance for information carriers.
## Repetition
Detection of changes in repetitive objects is poor.

## Incongruence
Effects are inconclusive. Relevant for road information, which is not always congruent with the road environment.

## Motion in Space
Motion in space does not solely induce change blindness, though interaction effects with other types of changes do exist. Only relevant for observer motion.

## Meaningfulness
Detection of changes in information of central interest is higher than of marginal interest. Very relevant. Drivers should consider information displayed to be relevant for the driving-task.

## Task-Dependency
Changes in task-dependant objects are more easily detected. Very relevant. Drivers should consider information displayed to be relevant for the driving-task.

## Expectancies
Detection of changes which are not expected or violate strong expectations, is poor. Very relevant.

Regarding the driving task, foveal information appears to be a confounding factor. In general it seems that changes in foveal information are likely to be detected, while more peripheral information like changes in objects on the sides of the road, are likely not to be detected. Results of Martens & Fox (2007) as well as Charlton & Starkey (2010) support this assertion. These findings are also supported by Uchida et al. (2011), who found drivers had difficulties with the detection of motion in the peripheral view. Galpin et al. (2009), on the contrary, found change detection of peripheral targets to be faster than centrally positioned targets. This finding could also be caused by using static pictures of road scenes, whereas both Martens & Fox’s (2007) and Charlton & Starkey’s (2010) participants were actually driving through the scene using a driving simulator. Using static instead of dynamic scenes might have reversed the visual scan pattern which is expected for visual search in driving, with a preference for more foveal objects (see Galpin et al. (2009) for reviews).

Change detection seems best for information that is integrated in the internal representation on a situation. The first pieces of information which are used to build this internal representation appear to come from schemata pointing out what is relevant information, where to find it and, depending on experience, what the information will be. Saliency, coarse object features and unique object existence also seem to be included in the first pieces used to build the internal picture, while spatial location does not.

What the different types of studies and their outcomes also point out, is that not only the task at hand, but also the stimuli and research paradigm used affect research findings.
4. CHANGE DETECTION OF CHANGES IN DYNAMIC SPEED LIMITS

The general objective of this literature review is to understand the effectiveness of dynamic speed limits in terms of perception. Will they suffer from change blindness? It has been established that information on information carriers competes with other (types of) information. To be perceived, this information must be selected as well as processed. As experience and the task at hand play a large role in detection, positioning dynamic speed limits where they are expected will enhance perception. Research of Hoogendoorn (2010) points out that compliance with a speed limit is attained when positioning them on gantries above the road, one each per driving lane. Although expectations concerning location can be met, the current literature review shows that change detection can still be an issue in using dynamic information for traffic management purposes.

Based on this literature review, it can be concluded that detection of changes in dynamic speed limits, will suffer from gap-contingency masking and changes in semantic identity. Placing dynamic speed limits on gantries imposes the problem of gap-contingency. Since there are a few hundred metres between gantries there are hardly any transient motion signals to be seen. As the change happens quickly and the gantry gets out of sight while driving, it is almost impossible for the transient motion signal of a speed limit change to capture attention. Detection of changes in dynamic speed limits may suffer from this. The change concerned is also a small semantical change, with the whole object remaining the same shape. Therefore object selection needs to be top-down. However, several studies have shown that drivers are not particularly interested in looking at traffic signs (see Charlton (2006) for reviews). Drivers who have seen the traffic sign over and over again shorten their glance duration while increasing the strength of their expectations about the semantic message of the sign. When strong expectations are violated, change blindness is increased (Martens & Fox, 2007).

As dynamic speed limits consist of road information, detection of changes might also suffer from repetition. It is unclear whether this repetition effect is only found for similar objects next to each other, or if it is also applicable for repetition in successive information carriers. Therefore the impact of this particular characteristic on change blindness remains uncertain. Dynamic speed limits also have characteristics favouring the likelihood of detecting changes in dynamic speed limits. As they are part of the driving task (maintaining speed), the chance of change detection increases. Drivers also have a preference for attending to foveal offered information, seemingly increasing the chance of change detection. This is also where the dynamic speed limits can be found. In addition to that, speed limits can be considered very relevant for the driving task at hand. Whether the dynamic speed limits are actually considered the center of interest by drivers remains to be examined. If they are, it could improve detection of changes.
In consideration of all the above, it is likely that some drivers will not detect changes in dynamic speed limits, but it remains unclear to what extent this will be the case. This literature review pointed out the specific factors which may play a role in detecting changes in dynamic speed limits. It shows further experimental research is needed to examine whether change blindness for changes in dynamic speed limits is due to the failing to look, or looking but failing to see. Respective questions are if dynamic speed limits are part of drivers’ center of interest, and whether drivers expectations about dynamic speed limits are met. Subsequent research should establish the extent of the change blindness effect and whether it can be reduced to increase situation awareness and the effects of dynamic traffic management.

**REFERENCE LIST**


