Stefano Manzo

DTU Transport, Technical University of Denmark,

E-mail address: stman@transport.dtu.dk

Investigating uncertainty in BPR formula parameters:

a case study

ABSTRACT

Transport models are subject to uncertainty, which refers to the impossibility of modelling with a deterministic approach. If not properly quantified, the uncertainty inherent in transport models makes analyses based on their output highly unreliable.

Within traffic assignment models, the relationship between travel time and traffic flows is commonly described by the BPR formula. Usually, the values for the BPR formula parameters are pre-defined based on assumptions and practice. The study described in this paper investigated uncertainty in the BPR parameters. Two dataset related to the Danish road network, namely Mastra and Hastrid, were analysed to estimate parameter values from observed data for three different types of roadways: highway, urban roads and local roads. BPR formula parameter distributions were generated by combining non-linear regressions with re-sampling Bootstrap technique. Latin hypercube sampling was then implemented on the results of this procedure and the generated parameter vectors were used to implement sensitivity tests on the four-stage model of the Danish town of Næstved.

The results clearly highlight the importance for modelling purposes of taking into account BPR formula parameters uncertainty, expressed as distribution of values, rather than assumed point values. Indeed, the model output demonstrates a high sensitivity to different parameter values and type of distribution. This proved true for all the three types of roadways analysed, highway, urban roads and local roads. However, different levels of parameters uncertainty, i.e. different levels of spread around the mean values, were observed for the different roadway classes.

Keywords: BPR formula, uncertainty, bootstrap, four-stage model

1 INTRODUCTION

By modelling complex systems, transport models are subject to uncertainty that can affect all model components (i.e. context, model structure and methodology, inputs and parameters) to finally propagate to the model output. The main consequence of this inherent uncertainty is that transport models do not provide reliable point estimates of modelled traffic flows and derived measures. Instead, modelled traffic flows are better expressed as a central estimate and an overall range of uncertainty margins articulated in terms of (output) values and likelihood of occurrence, as suggested by Boyce (1999). Uncertainty analysis relates to how uncertainty in each model component propagates to the model output and how to express the model output as a distribution, so reflecting the overall uncertainty present in the model.

Within traffic assignment models, the relationship between travel time and traffic flows is commonly described by the US Bureau of Public Roads (BPR) formula. The BPR formula works as a link performance function; given free flow travel time, observed flow and link capacity, it uses parameters to represent different relationships according to various types of roadways and circumstances. Usually, the values for the parameters are pre-defined, based on assumptions and practice. However, as for any other components of the assignment model, the BPR formula parameters have inherent uncertainty. With respect to BPR formula parameters, uncertainty originates from both the ignorance of the modeller of the true value of the parameters (epistemic uncertainty) and the stochastic behaviour of the (true) parameters itself (ontological uncertainty), which potentially vary by drivers behaviour, time of the day, weather conditions, link characteristics, etc.

In transport modelling uncertainty literature a common way to quantify model uncertainty is to run sensitivity tests on the model output by using inputs and parameter distributions, output of stochastic sampling procedures. For this purpose, with respect to model parameters uncertainty, random re-sampling techniques have been used, such as Jack knife in Armoogum (2003) and Bootstrap in Brundell-Freij (2000), Hugosson (2005), De Jong et al. (2007), Matas et al. (2011) and Petrik et al. (2012). The main advantage of these sampling approaches, as compared to the more frequently used Monte Carlo simulation, is that they do not require modellers' knowledge or assumptions on the parameter distributions shape. As argued in Hugusson, Bootstrap despite requiring more computations is likely to be a more efficient sampling method than Jack knife, if only because, unlike Jack knife, it does not assume the linearity of the parameter investigated. Bootstrap defines the parameter distributions by recalibrating the model parameters for a number of model samples, also called Bootstrap samples, which are generated from the original one by sampling with replacement.

On the top of our knowledge, no attempt has been made so far to estimate uncertainty in the BPR formula parameters from the analysis of observed data. For this purpose, we used observations from two datasets, Mastra and Hastrid, which refer to the Danish road network. Non-linear regression analyses, allowing calibrating the values of the BPR formula parameters simultaneously, were implemented and BPR formula parameters were estimated for three different road classes: Highway, urban roads and local roads. Parameters were repeatedly calibrated on 999 Bootstrap samples, as in Hugosson (2005) and Petrik et al. (2012), to generate parameter distributions. Latin Hypercube Sampling (LHS) procedure was then applied to create parameter vectors of 100 draws each which were used to run sensitivity tests on the transport model of the Danish town of Næstved. The use of LHS was required to reduce the number of model runs for the sensitivity tests, due to time constraints.

The following section provides the reader a description of the methodology applied to estimate the parameter distributions, including a description of the datasets used for the research and of the Bootstrap sampling technique. The section 3, after a brief description of the Næstved model, illustrates and discusses the results from the sensitivity tests run. The conclusions from this research are included in the last section of the paper.

2 METHODOLOGY

2.1 Speed-flow relationships

In transport assignment models the common way to describe the relationship between travel time and traffic flows is the BPR formula, as follows:

$$T_r = TF_r * \left\{ 1 + \alpha * \left[\frac{Flow_r + \gamma\left(Flow_r^{'}\right)}{Capacity_r} \right]^{\beta} \right\}$$

where T_r is the total travel time on link *r*, TF_r is the is the free flow time on link *r*, $Flow_r$ and $Capacity_r$ refer to the traffic volume and the capacity of link r, $Flow'_r$ refers to the traffic volume on the opposite direction of link *r* (relevant only in case of no separated lines), α and β are the traffic/delay parameters, and γ represents the effect on speed reduction due to opposite traffic in non-separated lane roads.

Suggested values for the formula parameters α and β vary according to the characteristics of the network object of the modelling exercise. The traditional values for α and β are 0.15 and 4 although higher values, respectively 0.84 and 5.5, have been suggested (Zhao and Kockelman 2001). For Danish road system, values for α have been observed between 0.8 and 1.2 and for β between 1.5 and 4, the larger the road the larger the β value (Nielsen and Jørgensen, 2008). Hansen (2011) still referring to the Danish road system suggests a broader range of values, between 0.5 and 2 for α, 1.4 and 11 for β and 0.05 and 0.2 for γ.

The available datasets, which will be illustrated in the section 2.2, do not include information about travel time. Thus, the BPR formula was modified to express the relationship between speed (instead that travel time) and flow/capacity ratio, such as in Nielsen and Jørgensen (2008) and Fagnant and Kockelman (2012), as follows:

$$S_r = \frac{SF_r}{1 + \alpha * \left[\frac{Flow_r + \gamma\left(Flow_r^{'}\right)}{Capacity_r} \right]^{\beta}}$$

where S_r is the observed average speed on link *r* and SF_r is the velocity in free flow condition in the link *r*. This transformed formula was used as specification for a non-linear regression model implemented to calibrate the BPR parameters using the statistical software SAS.

It is important to stress that this approach has two drawbacks. First, it implies an approximation. In fact, the speed is measured by local detectors so it does not reflect precisely

the link travel time, which is instead expression of the overall link conditions. On the top of our knowledge no attempt has been done so far to quantify this discrepancy. Second, the BPR formula proves to be correct to model travel time only when the traffic flow is below capacity. Indeed, when traffic flow reaches the capacity, in the figure 1 the point corresponding to flow at capacity (FC) and related speed at capacity (SC) the BPR formula curve takes the shape of the dotted curve on the right of FC. Instead, the observed traffic behaviour is tendentially close to the pattern described by the bold line. However, in static assignment models BPR formula is commonly used and accepted for practical reasons, among the others that in this way the speed flow relationship curve is "continuous even beyond capacity and differentiable" (Nielsen and Jørgensen, 2008).

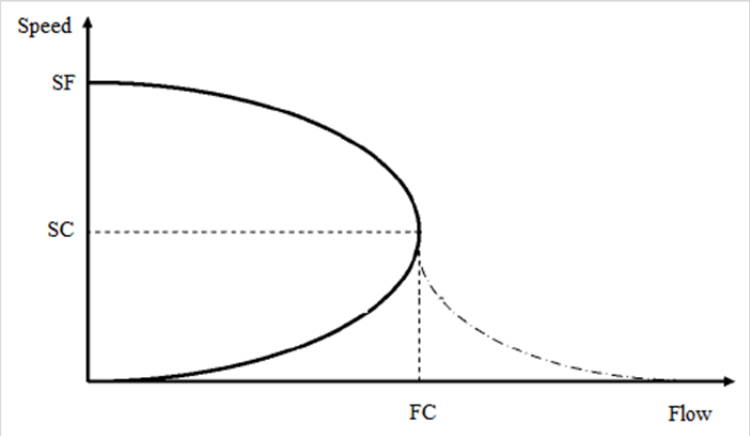


Fig. 1: Assumed relationship between speed and flow

2.2 Mastra and Hastri dataset

In this study we used two datasets, namely Mastra and Hastrid, both referring to the Danish road network. Mastra contains observed link capacity, flows and average speed by time intervals of 15 minutes, referring to three different types of roadways: highway, urban roads and local roads. Hastrid contains the same kind of observations but for highway network only. Table 1 summarises the main characteristics of the two datasets while figures 2, 3 and 4 graphically show average speed plotted against traffic flow observations by link type.

Table 1: Mastra and Hastrid datasets description

	Links			Observations			Capacity (average per hour)		
	Highway	Urban	Local	Highway	Urban	Local	Highway	Urban	Local
Mastra	22	6	6	12672	95520	3744	4275	1217	1576
Hastrid	11			3742			5443		

As can be seen from figures 2, 3 and 4, the datasets do not include many observations related to traffic in congestion conditions. Besides, especially with respect to urban and local roads, the datasets do not provide any observation for high flow traffic conditions. This, of course, proved to be a relevant problem with respect to the estimation of the BPR formula parameters. In fact, the parameters calibrated using the observations pooled by road class, summarised in table 2, resulted not always consistent with what expected, especially for highway and urban roads networks.

Table 2: BPR parameters (observations pooled by road class)

	Parameter	Estimate	StdErr	tValue	pValue
Highway	alpha	0.067	0.001	91.206	0.000
	beta	0.019	0.005	3.502	0.000
Urban	alpha	0.003	0.013	0.230	0.818
	beta	0.285	0.005	55.490	0.000
	gamma	343,389	5,254,340	0.065	0.948
Local	alpha	0.238	0.014	17.260	0.000
	beta	1.262	0.032	38.978	0.000
	gamma	0.189	0.055	3.468	0.001

For this reason for highway and urban roads it was implemented an analysis by link, to select the links providing results, in terms of calibrated parameters, within the range of values from the existing studies. Based on the results of this analysis, 7 and 3 links were selected for, respectively, highway and urban roads; a new calibration was then implemented on these restricted datasets providing the values summarised in table 3.

Table 3: BPR parameters (selected Highway and urban road links)

	Parameter	Estimate	StdErr	tValue	pValue
Highway	alpha	0.672	0.053	12.755	0.000
	beta	5.510	0.269	20.504	0.000
Urban	alpha	0.166	0.005	35.661	0.000
	beta	0.585	0.008	77.571	0.000
	gamma	0.646	0.076	8.469	0.000

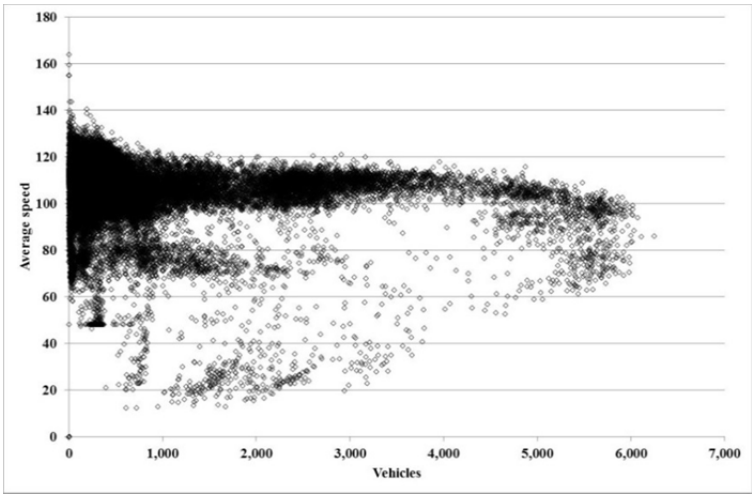


Fig. 2: Mastra and Hastrid highway speed-flow observations

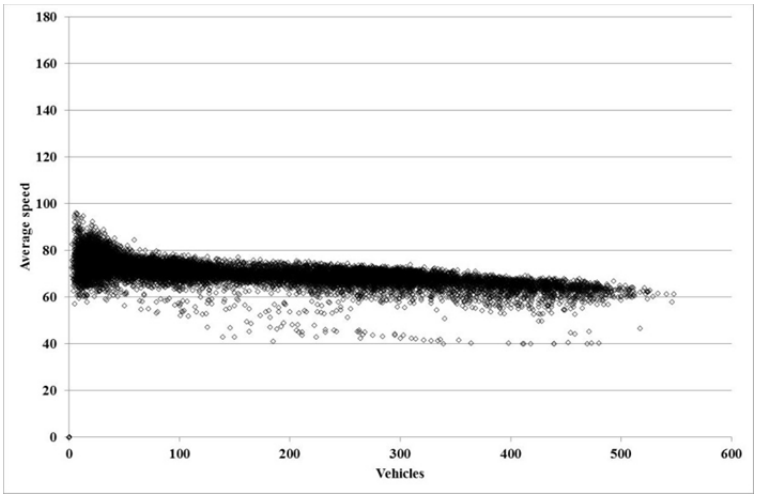


Fig. 3: Mastra urban roads speed-flow observations

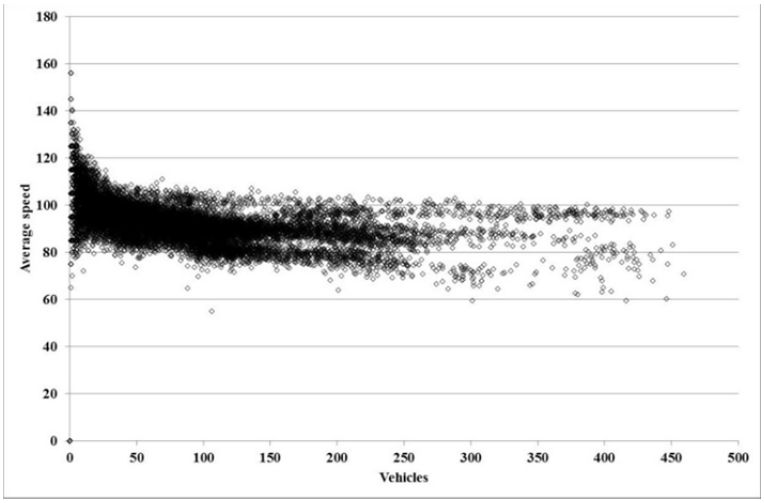


Fig. 4: Mastra local roads speed-flow observations

2.3 Bootstrap analysis

In order to produce BPR parameter distributions, the re-sampling technique Bootstrap was used. Bootstrap method investigates the accuracy of an estimator θ starting from the initial assumption of considering the original sample, originating θ, as the population. Bootstrap consists in a three steps procedure. First, from the original sample of *n* observations a number of samples are generated through (re)sampling with replacement, all samples containing *n* observations as the original one. The replacement approach guarantees that each observation in the original sample has a constant probability 1/n to be drawn and placed in the new generated samples so that these, also called Bootstrap samples, differ from each other. Second, the estimator θ is calculated for each of the new generated samples. Finally, the new θ values obtained are analysed to infer on the accuracy of the estimator by using one of the numerous uncertainty measures available from the literature (e.g. variance, standard deviation, confidence intervals or percentiles).

It is important to notice that the Bootstrap method has two downsides. First, there is no rule defining the correct number of Bootstrap samples to generate, although the number should be large and in theory tendentially infinite. Second, the results are constrained by the quality of the original sample, given that the Bootstrap samples do not increase the amount of information there contained.

Using as original samples the selected link samples for Highway and urban roads and the full sample for local roads, 999 Bootstrap samples were created and the calibration process was repeatedly implemented for each of them. The resulting parameter statistics and fitted distributions, based on Kolmogorov-Smirnov (K-S) test, are summarised in table 4. Also Coefficient of Variation (CV) are reported and henceforward used as measure of uncertainty. As can be noticed, the highest level of uncertainty is related to the values of the gamma parameter, for both urban and local roads. Also the alpha parameter for the Highway network shows a high level of CV. Interestingly, different parameters of different road classes resulted to have different distributions.

Table 4: Bootstrap parameters statistics

	Parameter	Estimate	StDev	Min	Max	CV	K-S
Highway	alpha	0.675	0.079	0.450	0.984	0.118	Lognormal
	beta	5.510	0.385	4.246	6.796	0.065	Normal
Urban	alpha	0.166	0.006	0.149	0.183	0.035	Normal
	beta	0.585	0.007	0.564	0.610	0.012	BetaGeneral
	gamma	0.651	0.093	0.418	0.970	0.144	Lognormal
Local	alpha	0.237	0.011	0.205	0.284	0.046	Normal
	beta	1.261	0.015	1.212	1.311	0.012	InvGauss
	gamma	0.193	0.038	0.081	0.328	0.197	Gamma

3 CASE STUDY

3.1 The Næstved model

The uncertainty analysis was implemented on a four-stage transport model of the Danish town of Næstved. The four-stage transport model is an analytical framework that combines trip generation, trip distribution, mode choice and trip assignment (see, e.g., Ortuzar and Willumsen, 2011). Each model output is used as input for the model that follows, and the link flows from the trip assignment are used as feedback for the previous stages of the framework. The model is solved with an iterative procedures that concludes when the link flows reach equilibrium, which usually corresponds to either the deterministic or the stochastic user equilibrium state (for details see, e.g., Ortuzar and Willumsen, 2011). Given the wide use of the four-stage transport model framework, results from this study are straightforward to interpret and to compare with other literature and project results.

Næstved is a Danish town located in the southern part of Zealand, with a population of around 42,000 increasing to around 80,000 when considering the entire municipality, which has a total surface of around 681km. The total number of trips over a 24h time interval is estimated of around 88,500, 10% of which made by public transport through a network of buses connecting Næstved to its urban area, as well as all major surrounding towns. In the Næstved model, the area of interest is divided in 106 zones. The network is composed by links classified as "highway", "urban" and "local" which represent respectively around 3%, 5% and 92% of the total number of links. The traffic, modelled over a single 24 hour time interval, is divided in two modes, private and public transport, and in two categories, home/work and business. The model final output is based on 3 model's iterations which only involve trip distribution, mode choice and trip assignment stages; in other words the trip generation output is kept constant and is not influenced by the travel impedance of the network. Due to the small size of the town of Næstved, the network is characterized by low levels of congestion.

For this study trip generation, trip distribution and mode choice were considered uncertainty free, whilst instead the focus was on the assignment model that is a link-based model solved by the Method of Successive Averages (MSA) to reach Stochastic User Equilibrium (SUE). The chosen route to travel by mode *k* between zones *i* and *j* is the one that minimizes the cost of travelling, estimated at the link level and calculated as:

$$c_{ijkr} = \omega_l L_{ijkr} + \omega_{tf} TF_{ijkr} + \omega_{tc} TC_{ijkr} + \varepsilon_{ijkr}$$

where c_{ijkr} is the cost of travelling by mode *k* from zone *i* to zone *j* using link *r*, L_{ijkr} is the length of the link *r* by mode *k* from zone *i* to zone *j*, TF_{ijkr} is the free flow travel time, TC_{ijkr} is the extra travel time due to congestion, ε_{ijkr} is the vector of residuals, and the ω's are the

parameters associated to the respective variable. The travel time/flow relationship is based on the BPR formula.

To investigate on model output sensitivity to BPR parameters uncertainty, in order to reduce the number of model runs due to time constraints, LHS procedure was applied on the (Bootstrap) parameter distributions to create parameter vectors of 100 draws. Each of 100 draws parameter combinations was then used to run sensitivity tests on the Næstved model. Table 5 below summarises the results from the LHS, so providing information about the BPR formula parameter values variation domain used in the sensitivity tests.

Table 5: LHS statistics

	Parameter	Mean	StDev	Min	Max	CV
Highway	alpha	0.675	0.081	0.457	0.928	0.120
	beta	5.508	0.357	4.529	6.353	0.065
Urban	alpha	0.166	0.006	0.153	0.182	0.035
	beta	0.585	0.007	0.565	0.606	0.012
	gamma	0.651	0.093	0.459	0.916	0.143
Local	alpha	0.237	0.011	0.209	0.267	0.046
	beta	1.261	0.015	1.223	1.303	0.012
	gamma	0.193	0.037	0.111	0.290	0.194

3.2 Results and discussion

The results from the sensitivity tests run on the Næstved model are summarised in tables 5, 6 and 7.

Table 6: Vehicle-kilometre (link)

	CV
Total	0.127
Highway	0.040
Urban	0.249
Local	0.122

Results shown in table 5, which refer to the average vehicle-kilometer CV calculated at link level for the overall network and by road class, demonstrate a relevant sensitivity of the model output to the BPR parameters uncertainty, with a CV for all the links of 0.127. It is worth to notice that, by running the model with the SUE approach but assuming no uncertainty in the BPR parameters, the model uncertainty is equal to 0.059. Urban road links show the highest level of uncertainty, followed by local links, probably due to the higher number of route choice alternatives that both networks offer as compared to the highway network.

Table 7: Vehicle-kilometre

	Mean	St Dev	CV	Distribution
Total	2,737,578	2,415	0.001	Gamma
Highway	694,335	15,320	0.022	Logistic
Urban	411,553	11,469	0.028	Loglogistic
Local	1,631,690	8,832	0.005	Logistic

Table 8: Network travel resistance

	Mean	St Dev	CV
Free time	2,754,855	4,391	0.001
Cong time	37,048	4,818	0.130

As shown in table 6, the uncertainty related to the overall amount of vehicle-kilometre output is small, with a CV of 0.001. This is probably due to the low levels of congestion in the network. In fact, if in one hand a small variation of number of trips may cause a noticeable change in vehicle-kilometre at link based level, as discussed above, on the other hand in terms of total amount of traffic it does not. However, also in this case different road classes have different sensitivity to BPR parameters uncertainty, with urban roads and highway showing a similar and higher CV as compared to local roads. The sensitivity tests also demonstrated a relevant sensitivity of the model in terms of modelled congested time whose CV, as shown in table 7, was equal to 0.130.

4 CONCLUSIONS

The model output analyzed were (i) vehicle-kilometre on single links and overall network and (ii) travel resistance within the network. The results clearly highlight the importance for modelling purposes of taking into account BPR formula parameters uncertainty, expressed as distribution of values, rather than assumed point values. Indeed, the model output demonstrates a high sensitivity to different parameter values and type of distribution.

Moreover, different road classes have shown different sensitivity to BPR parameters uncertainty. This seems to suggest the possibility of developing a class reference approach for uncertainty analyses of such kind, so advising further research on the topic.

The speed-flow data analysis produced different parameter distributions with respect to the three different road classes. These is a result of relevant interest because it reaffirms the importance, within sampling procedures, of defining parameter distributions (and mean values) from observed data rather than infer them from assumptions, whenever it is possible. Indeed, the parameter distributions (and mean values) for specific networks may significantly vary from the standard ones suggested by the existing literature or imported from previous studies. Thus, pre-defined parameter distributions (and mean values) based on assumptions and practice may not be reliable for modelling purposes.

REFERENCES

Akcelik, R. (2003). Speed-flow models for uninterrupted traffic facilities. Technical report.

Brundell-Freij, K. (2000). Sampling, specification and estimation as sources of inaccuracy in complex transport models - some examples analysed by Monte Carlo simulation and bootstrap. *Proceedings of the European Transport Conference,* Cambridge (UK), 2000.

De Jong, G., Daly, A., Pieters, M., Miller, S., Plasmeijer, R., Hofman, F. (2007). Uncertainty in traffic forecasts: literature review and new results for The Netherlands. *Transportation*, **34**, 375-395.

Fagnant, D.J., Kockelman, K.M. (2012). Outcome of transportation projects under uncertainty. *Transportation research records*, **2303**, 89-98.

Hansen, C. (2011). Estimation af nye speed-flow kurver. Technical note (unpublished results).

Hugosson, M.B. (2005). Quantifying uncertainty in a national forecasting model. *Transportation Research Part A*, **39**, 541-547.

Matas, A., Raymond, J., Ruiz, A. (2012). Traffic forecasts under uncertainty and capacity constraints. *Transportation*, **39**, 1-17.

Nielsen, O.A., Jørgensen, R.M. (2008). Estimation of speed-flow and flow-density relations on the motorway network in the greater Copenhagen region. *IET Intelligent Transport Systems*, **2**, 2, 120-131.

Ortuzar, J., Willumsen, L.G. (2011). Modelling transport 4th edition. John Wiley & Son Ltd, Chichester, United Kingdom.

Petrik, O., Mercier Vilaca e Moura, F.M., Ghebrehiwot, Y., de Abreu e Silva, J. (2012). Measuring uncertainty on the Portuguese high-speed railway demand forecast models. *Proceedings of the 13th International conference on travel behaviour research*, Toronto, July 2012.

Zhao, Y., Kockelman, M. (2001). The propagation of uncertainty through travel demand models; an explanatory analysis. *Annals of Regional Science*, **36**, 145-163.