Modelling versus simulation: two approaches for the assessment of the service quality – cost relationship in innovative public transport systems*

Marco Diana
Dipartimento di Idraulica, Trasporti e Infrastrutture Civili
Politecnico di Torino
Corso Duca degli Abruzzi, 24
Torino, Italy I-10129
Tel. +39 011 564 5605 – Fax +39 011 564 5699
E-mail marco.diana@polito.it

Abstract

Among the "paratransit options" initially made possible by the development and availability of computers, Demand Responsive Transit Services (DRTS) were in the past regarded as a promising way of developing a more efficient public transport service. However the performances of the systems that were widely implemented from the 70s onwards were generally disappointing. Nowadays DRTS are experiencing a technology-driven renaissance, that will necessarily have to face the problems that came out in former systems: ridership overestimation, heavy operational costs, difficulties in managing situations that have not been foreseen in the planning phase. The core problem is to estimate how costs will be affected by the quality level of the offered service (waiting time at pickup points, extra ride time etc.). In this first phase of the research issues concerning demand estimation have been disregarded and we applied two different methodologies, namely a continuous approximation model and a combinatorial optimisation heuristic, in order to forecast the number of vehicles needed to operate the service as quality requirements become more stringent, for various demand levels. The application of two different methodologies for solving the same problem allows some interesting insights concerning benefits and drawbacks of each one. In particular, the modeling approach is much quicker and requires less data, but more research is needed to adapt it under a wider range of cases. On the other hand, the simulation approach is still the best choice when a more complete economic analysis is needed.

Conference* research domain

Transport economics

* Paper presented at the Joint ECTRI-FERSI Young Researchers Seminar, Lyon-Bron, 16th-18th December 2003.

1. Problem statement

Demand Responsive Transit Services (DRTS) are a particular form of public transport characterized by the fact that the vehicles (cars, vans or minibuses) operate in response to calls from passengers to the transit operator, who then dispatches a vehicle to collect the clients and transport them to their destinations. Unlike taxicabs, the operator tries to contemporarily serve more than one request with the same trip; as a consequence, passengers do not directly travel from origin to destination.

Among the "paratransit options" initially made possible by the development and availability of computers, DRTS were in the past regarded as a promising way of building a more efficient public transport service. After a decade of tumultuous development of these systems across the U.S. and abroad, there has been at the beginning of the Eighties a general disappointment for their economic behavior, which generally has led to their demission or radical transformation. A very often mentioned ex-post justification pinpoints the fact that the technological level was in these years clearly insufficient and informatics tools very expensive. Capitalizing on this argument, DRTS are recently experiencing a renaissance that from academic communities is now reaching the implementation level.

However we believe that there is still a fundamental passage that must be kept into consideration. Given the intrinsic complexity of the organizational form of any paratransit service, a more thorough economic analysis must be exploited in order to possibly avoid past errors. If we look back at the situation in the Seventies, it can be seen that most projects were based on over-optimistic assumptions concerning both the patronage and the cost structure in the long run. Here we would like to focus on methodologies that can be helpful in clarifying the key point that must be considered in the planning phase of a DRTS: the relationship between costs and quality of the service. It is anticipated that a fuller economic analysis of a DRTS would require many more research efforts that are beyond our scope.

Whenever we consider an industrial process aimed at producing goods or services, we know that it is possible to fix different cost-quality equilibria. For example, in the traditional transit industry a greater comfort onboard the vehicle can lead to purchase buses that are more expensive or have less seats, or decreasing the waiting times on a fixed-line system implies using more vehicles. The economic impact of these options can usually be straightforwardly determined, so that the cost-quality relationship can be easily deducted in these cases. However things are not so simple when we consider a DRTS, where the link between input (the

resources we need) and output (the service we provide) is not so clear, essentially because it is very difficult to express it in analytic form. This is essentially due to the fact that the operation of the service and hence its costs are *responsive*, i.e. completely dependent of the service demand, so that we cannot fix it like in traditional transit services. The link between demand and offer can be described through a combinatorial optimization problem, known as Pickup and Delivery Problem (PDP) or Dial-a-Ride Problem (DARP). The particular complexity of such problem has stimulated a great body of research (for its review, see for example Savelsbergh and Sol, 1995 and Desaulniers et al., 2000), and prevent us from easily understanding the relationship we are interested in investigating.

Considering the past story of paratransit systems, we believe that this incertitude concerning the cost-quality trade-off has been at least as much harmful as the often mentioned technological lack. It is thus very important to build tools that might be of help for decision makers that have to plan a new service. Given the broadness of the theme, in the following we will focus our attention on a specific issue, i.e. the determination of the number of vehicles needed to provide a service of pre-determined quality. In Diana (2003) two possible methodologies have been defined in order to achieve this result, namely the utilization of a scheduling algorithm that can simulate services of different quality and the application of a continuous approximation model. Our goal is now to draw a comparison between these complementary approaches, in order to characterize their benefits, drawbacks and ambits of use. The number of vehicles needed to operate the service is linked to the number of drivers on duty, so that it determines most of the costs. There is of course a component that is linked to the traveled length, but it will not be the main concern in this paper.

The remainder of the paper is organized as follows. In section 2 we exactly define the problem we are going to study and we present the simulation approach, whereas in section 3 the concurrent continuous approximation model is shown. Then we present the results of the two approaches over a given experimental plan and we finally perform a comparison between these two solution techniques.

2. Definition of the envisioned DRT system and simulation approach

In the following we will adopt a simplified version of the operating scenario described by Jaw et al. (1986). Our DRTS consists of a fleet of vehicles with no predefined schedule. The vehicles travel at a constant speed and cannot idle. The service time at the nodes is zero and we do not consider capacity constraints, since in all the practical cases they are dominated by time window constraints. When making a reservation, the customer has to specify the origin and the

destination of the trip, as well the pickup time. In order to ensure an acceptable quality of the service, the vehicle has to pickup the customer no later than a predetermined time interval from the specified pickup time, and the maximum length of the trip must be somewhat limited. To do this, we fix a maximum wait state, that is the same for all the customers, and we compute a maximum ride time for each request. The maximum ride time is defined as an increasing function of the direct ride time, that is of the time needed to serve the request without deviations. It is convenient to merge these constraint, related to the quality of the service to be provided, into the definition of the time windows for all the pickup and delivery nodes. Details concerning the time windows computation can be found in Diana and Dessouky (2003).

Now, we would like to understand how many vehicles are needed to provide the described service when we have a set of requests, to be served using a known road network, with specific quality requirements expressed in terms of time window size of the pickup and delivery points. In order to do this, we schedule the service by using the parallel regret insertion heuristic that has been presented in Diana and Dessouky (2003). The input data of the algorithm are the list of requests, the travel times and distances between all the pairs of service points, the number of vehicles to be used and their capacity (number of seats), the depot location and the time windows. The algorithm provides the schedule of the service, and a separate list of the requests that have been rejected, i.e. that cannot be serviced without a time windows violation. If this list is not empty, the solution is not feasible and we run again the simulation by incrementing the fleet size, until all the requests are served. In this way, the minimum number of needed vehicles is determined.

In order to understand the relationship between number of needed vehicles and time windows size, we can run several different simulations, by changing the quality requirements (i.e., wait state and maximum ride time). It is thus possible to see, given a set of requests, that the higher the quality, the larger the number of vehicles needed to find a feasible solution. This is an intuitive result, but the proposed methodology allows to quantitatively define the trade-off between the two elements. Of course the result is valid only for that particular set of requests in the given service area, so that several sensitivity analyses (and a potentially very high number of simulations) are needed if we are looking at more general results.

3. A continuous approximation model for the fleet dimensioning problem

From the description of the simulation approach that we reported in the previous paragraph, it can be seen that the most important output of the heuristic, that is the schedule of the service, is not needed for our purposes. In fact, we are only interested in knowing whether all the

requests can be served with a given fleet dimension. Hence it would be interesting to find a less computationally burdensome approach, that can directly predict the number of needed vehicles for a specific spatial and temporal demand pattern in a service area, without having to determine in details the service operation.

This result is usually achieved through the utilisation of a continuous approximation model, i.e. a model that does not consider the list of requests but looks at the travel demand in a more aggregate way, through its spatial and temporal density distribution. Such approach is generally being used for problems that are simpler than ours, so that they can usually be tackled relying on the well known approximation of the length of a Traveling Salesman tour (Beardwood et al., 1959; Eilon et al., 1971), or other Geometrical Probability methodologies. However, when there are time windows, that are the key point in our investigation since they are related to the quality of the service, these results cannot generally be applied. To the best of our knowledge, the only model that addresses this issue is the one of Diana et al. (2003). This model takes as input the mean travel speed of the vehicles, the spatial and temporal distribution of the travel demand and the time windows width, and directly gives an estimation of the number of needed vehicles. Generally speaking, this is done by computing the joint probability of serving with one vehicle the sequence of pickup and delivery nodes. A full description of the model can be found in the mentioned reference.

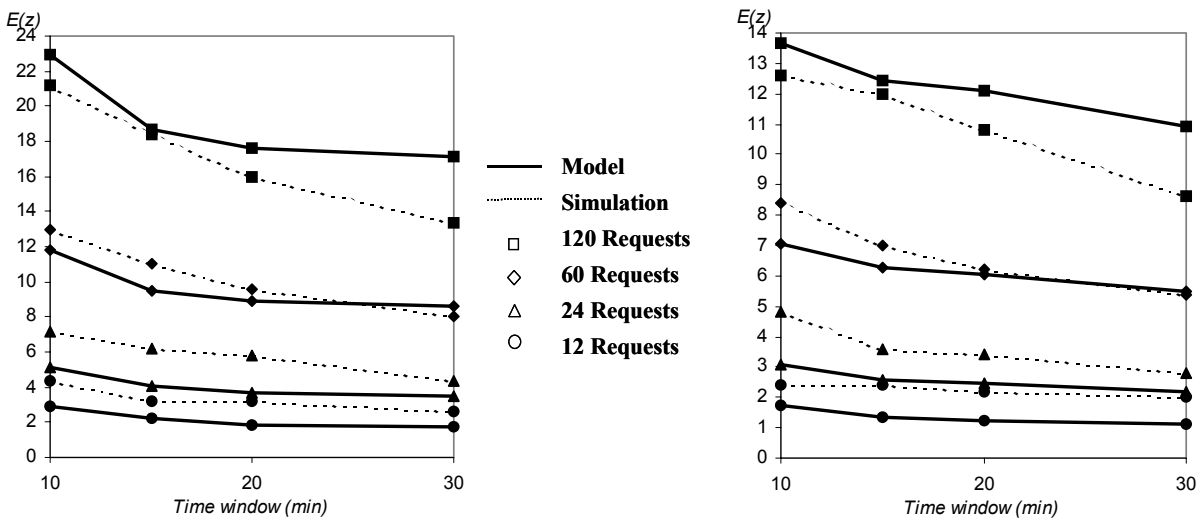
This model is still under investigation, in order to test its reliability under a wide range of cases (demand levels and time windows). Preliminary computational tests have shown its fairly good behaviour when the demand density is not too high. We refer the interested reader to the Diana et al. (2003) paper for a more complete description of these results.

4. Computational experience with the two methodologies

In order to carry out a thorough comparison of the two approaches, that is the aim of this paper, we have to define a common experimental plan. We consider a square area of 10*10 miles and a planning period of 2 hours. We use a short planning period since we are focusing on determining the fleet size during the peak period. We assumed that pickup and delivery points are either uniformly or normally distributed over the square area. In the latter case, the distribution is centred in the middle of the square and its variance is 100/36 square miles. In both cases we varied the number of requests over the planning period from 12 to 120, the requests arrival being a Poisson process. In order to apply the simulation, we generated 5 samples of requests drawn from these statistical distributions for each considered scenario, whereas the mathematical expressions of the distributions themselves are used by the model.

Finally, we associate to each pickup and delivery node the same time window, ranging from 10 to 30 minutes.

Computational results in terms of expected number of vehicles *E(z)* are reported in figure 1. The continuous lines represent the outcome of the approximation model, whereas the dotted ones report the results from the simulation approach. It can be seen that there is fairly good correspondence between the two alternative methodologies, so that they could be both used for practical purposes. This is particularly true when the rideshare probabilities are not so high; except the case of 120 requests served with the largest time window, difference of *E(z)* from the two methods is always below 3 vehicles.



(a) (b)

Figure 1. Fleet dimension as the number of requests and the time window changes for uniformly (a) and normally (b) distributed requests

As expected, the plots indicate that when we increase the number of request and we narrow the time windows, we need more vehicles in order to schedule the service. It is however interesting to note that the quality-cost and demand level-cost relationships are not linear, due to the rideshare mechanism. In other words, when we increase the service level in terms of time windows width, the costs increase more than proportionally. For example, considering the simulation results for a set of 120 uniformly distributed requests, when we narrow the time windows from 30 to 20 minutes we have to use 2 more vehicles (from 14 to 16), whereas from 20 to 10 minutes we need to add 5 further vehicles. On the other hand, for example halving the patronage does not consent to halve the fleet. This behaviour is essentially due to the fact that the number of needed vehicles is linked to the rideshare, that is the probability of serving more than one request with a single trip. This probability is of course not linearly increasing

with the number of requests and the width of the time windows, but is ruled by much more complicate mathematical laws that have been investigated in Diana et al. (2003).

To sum up, we believe that the reported plots are a useful tool in the planning phase in order to help fixing the right compromise between quality targets and needed resources.

5. Comparison of the two methodologies

Since we presented two alternative tools for determining the same result, it is now interesting to see benefits and drawbacks of each one, in order to determine their optimal range of application. A good starting point for a theoretical discussion can for example be represented by the analysis carried out by Larson and Odoni (1981), where the simulation approach is critically reviewed in comparison with idealized models. As these authors point out, when we do not need too detailed results, there are problems in which the simulation is almost useless, since an easy mathematical expression can be derived. In other cases, mathematical results are not available or not reliable or a more detailed description of the phenomenon is sought. Out of the Diana et al. (2003) paper, the problem of the quality-cost relationship in a many-to-many DRTS service with time windows has only been tackled through a simulation approach, due to the difficulties of an analytic approach when time windows must be respected. The mentioned paper presents a continuous approximation model, but the research on this topic is still at an early stage and the mentioned results are not sufficient to conclude that the simulation approach is no more useful, even for more coarse-grained analyses.

On the basis of the computational experience that has been presented in the previous section, considering also the input data and the computational effort needed in each method, we can say that there is not a procedure that clearly outperforms the other. The simulation approach gives results that allow for a better estimation of the quality-cost relationship, since the definition of the schedule implies the possibility of keeping into account more details, such as the travelled distance, the crew roster design, the influence of the depot location etc. On the other hand, it is necessary to run a specific simulation for every scenario we want to test, and moreover we need to know very detailed data concerning the list of requests and the network characteristics. If these input data are not reliable, much of the precision of the output is of course an illusion. The approximation model provides a more rough estimation of the service costs, since it computes only the number of vehicles. However its computational burden is a fraction of that of a simulation, so that it is quite easy to explore a wide range of scenarios. Also the input data are easier to estimate, since we are interested in knowing only the mean

spatial and temporal distribution of the demand, as well the mean commercial speed of the vehicles.

On the basis of these considerations, it is quite evident what are the best ambits of use of each tool. The approximation model is particularly indicated in those situations in which the customer behaviour is not known, for example when we have to implement a new service from scratch. On the other hand, at a preliminary stage of the planning process the key issue is the determination of the number of vehicles that must be bought, that is the most significant financial effort. In such a phase it would be quite hard to predict the exact location of all the requested service points, and therefore the simulation approach would not be very reliable. The utilization of a model can give the planner an overview of the quality-cost relationship, so that the best choice (for example in terms of kind of vehicles) can be done.

On the contrary, the simulation is surely recommendable whenever we have sufficiently detailed data and we want to test only a limited range of scenarios. A typical example is the necessity of forecasting the behaviour of an existing service when we modify some operating condition (changing the service area, improving the quality etc.). In those cases the effects over the operating costs cannot easily be captured with the presented approximation model.

Of course there is a wide range of cases that are somehow between these two extremes. For example, it is possible that we need to verify more thoroughly the behaviour of the system after the preliminary planning phase, once we have set up the fleet and the centre and the community is aware that "something is going on". Thus, it could be possible to punctually estimate the service demand, for a given level of service, through an accurate SP survey that would take into consideration also the substitution and complementary effects with existing transportation modes. Finally, it would be possible to run one or a few simulations on the basis of these data.

To sum up, in most of real-life situations a combined use of the two approaches could be envisaged, since they are somehow complementary, even if they both investigate the behaviour of a DRTS considering the supply side only. A project aimed at implementing a large-scale DRTS for the city of Turin has been launched, and it will constitute the occasion for the full implementation of the methodology that has been described here, from the preliminary feasibility evaluation till the monitoring of the service operation during its experimental phase.

References

Beardwood, J., Halton, J.H. and Hammersley, J.M. (1959) The shortest path through many points. *Proceedings of the Cambridge Philosophical Society*, *55*, 299-327.

Desaulniers, G., Desrosiers, J., Erdmann, A., Solomon, M.M. and Soumis, F. (2000) *The VRP with pickup and delivery*. Cahiers du GERAD G-2000-25, Ecole des Hautes Etudes Commerciales, Montréal.

Diana, M. (2003) *Methodologies for the tactical and strategic design of large-scale advance-request and real-time demand responsive transit services*, Ph.D. Dissertation, Politecnico di Torino, Dipartimento di Idraulica, Trasporti e Infrastrutture Civili, Torino, Italy.

Diana, M. and Dessouky, M.M. (2003) A new regret insertion heuristic for solving large-scale dial-a-ride problems with time windows. *Transportation Research B*, in press.

Diana, M., Dessouky, M.M. and Xia, N. (2003) A continuous approximation model for the fleet dimensioning of many-to-many demand responsive transit services with time windows. Submitted to *Transportation Science*.

Eilon, S., Watson-Gandy, C. and Christofides, N. (1971) Expected distances in distribution problems. In *Distribution Management: Mathematical Modelling and Practical Analysis*. Griffin, London, 151-179.

Jaw, J.J., Odoni, A.R., Psaraftis, H.N. and Wilson, N.H.M. (1986) A heuristic algorithm for the multi-vehicle many-to-many advance request dial-a-ride problem. *Transportation Research B*, 20B, 243-257.

Larson, R.C. and Odoni, A.R. (1981) Advantages, disadvantages and misuse of simulation. In *Urban Operations Research*. Prentice-Hall, Englewood Cliffs, 521-524.

Savelsbergh, M.W.P. and Sol, M. (1995) The general pickup and delivery problem. *Transportation Science*, 29, 17-29.