

**A MODEL FOR THE FLEET SIZING OF DEMAND RESPONSIVE
TRANSPORTATION SERVICES WITH TIME WINDOWS**

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 diana@polito.it

Maged M. Dessouky*

Daniel J. Epstein Department of Industrial and Systems Engineering

University of Southern California

3715 McClintock Avenue

Los Angeles, CA 90089-0193, USA

Tel. +1 213 740 4891 – Fax +1 213 740 1120

E-mail maged@usc.edu

Nan Xia

Department of Mathematics

University of Southern California

1042 Downey Way

Los Angeles, CA 90089-1113, USA

Tel. +1 213 740 4891 – Fax +1 213 740 1120

E-mail nxia@usc.edu

* Corresponding Author

ABSTRACT

We study the problem of determining the number of vehicles needed to provide a demand responsive transit service with a predetermined quality for the user in terms of waiting time at the stops and maximum allowed detour. We propose a probabilistic model that requires only the knowledge of the distribution of the demand over the service area and the quality of the service in terms of time windows associated of pickup and delivery nodes. This methodology can be much more effective and straightforward compared to a simulation approach whenever detailed data on demand patterns are not available. Computational results under a fairly broad range of test problems show that our model can provide an estimation of the required size of the fleet in several different scenarios.

KEYWORDS

Public transportation; Probabilistic models; Continuous approximation models; Paratransit services; Demand responsive transit systems.

1. INTRODUCTION

Demand responsive transit (DRT) systems are a form of flexible public transportation service in which the itineraries and the schedules of the vehicles are programmed on the basis of the requests of the users. Early DRT systems were designed for the general public, but within some years they met with financial problems and had to be discontinued or radically transformed (Lave et al., 1996). Nowadays, excluding some flourishing niche services such as airport feeders, existing DRT systems are almost exclusively used for dedicated services in eligible categories (e.g. disabled and elderly) and are heavily subsidized (Palmer, Dessouky, and Abdelmaguid, 2004).

In recent years there has been an increasing interest in reconsidering the implementation of these services in a broader set of circumstances. Technological advances can now provide Intelligent Transportation System devices, such as Automatic Vehicle Location or smart cards for fare collection, at a low cost. A large amount of research work has been carried out concerning the design of more efficient scheduling and routing algorithms of such systems; two of the latest comprehensive reviews can be found in Savelsbergh and Sol (1995) and in Desaulniers et al. (2001). Improvements in routing and scheduling may help overcome the economical inefficiency that can be usually detected in existing systems. In this case, a “smart” DRT system can be seen as one of the possible alternatives to offer a service of acceptable quality when the travel demand density is too low to justify fixed route lines.

To make this a feasible option, a set of transportation modeling tools is needed to effectively plan the system. Up to now, research efforts have mainly focused on the improvement of different aspects related to the service scheduling in order to achieve better efficiencies for the existing systems. When considering the tactical and operational decision levels, detailed planning activities related to the operation of the service must be carried out typically using mathematical programming techniques. However, during the design phase, the extensive datasets needed to perform such analyses are usually unavailable. Furthermore, as the system becomes bigger and more complex, the computational burden associated with efficient algorithmic procedures increases very rapidly. In those cases, the use of approximation models to estimate the costs on

the basis of a few inputs may be sought. The utility of this kind of technique has already been shown in related research fields, such as in logistic systems (Robusté et al., 1990; Daganzo, 1991; Langevin et al., 1996) and for the heuristic solution of the Traveling Salesman Problem (Del Castillo, 1999).

In this paper we propose an approach to address one of the critical aspects that must be considered on a strategic planning level, that is the determination of the required fleet size for a DRT system to meet a given demand level with a predetermined level of service. As will be later explained, we define the service level by a time window length specifying the maximum deviation from the requested pickup time and the maximum deviation over the direct ride time allowing for detours to service other passengers. Quadrifoglio, Dessouky, and Ordóñez (2005) use a scheduling algorithm embedded in simulation model to study the impact of the time window length on the fleet size and total trip miles driven by the vehicles. The intent of this study is to develop an analytical modeling approach as opposed to a simulation model.

Unlike standard optimization procedures that require the knowledge of the exact spatial and temporal location of the demand points, we propose a methodology in which only the spatial and temporal distribution of demand are known. Our methodology makes some simplifying assumptions regarding these distributions. We first develop the general model and then introduce the simplifying assumptions one by one. We show that these assumptions are not too restrictive by comparing our results to the results from simulation.

In the following section a review of the existing research activities in related fields is presented. Then the studied DRT system is defined. Section 4 presents the model and Appendix 1 shows its possible variants under various simplifying assumptions. Section 5 reports some computational experiments that show the effectiveness of our approach, whereas Appendix 2 is devoted to the presentation of former research works relevant to the numerical approximation we propose in our tests. Finally we discuss the results and suggest directions for future research.

2. LITERATURE REVIEW ON RELATED FIELDS

A good deal of research work has been devoted to the investigation of issues similar to ours, and it is interesting to compare different approaches in order to draw lessons for our problem. As will be shown in the following sections, the proposed methodology relies on a variety of research fields. In the following, we will briefly review them, pointing out the common features as well as the differences with our approach.

One of the related fields that can provide useful insights for our problem is the abundant literature on the modelling of freight distribution systems. The interested reader is referred to Langevin et al. (1996) for a more complete review. Here it is however interesting to point out that many of these models are based on a handful of theoretical papers concerning the expected length of a Travelling Salesman Problem (TSP) tour. The first work in providing an approximate formula is due to Bearwood et al. (1959). Eilon et al. (1971) estimate the length of a TSP tour for a fleet of vehicles through simulation. In distribution problems it is usually possible to divide the service area into several zones each of which is served with one vehicle and each path is estimated using the formulas for the TSP tour. This very popular technique, usually called “cluster-first, route-second”, has been successively used in many papers modelling distribution problems. They can take into consideration irregular areas and the most efficient shapes of the zones for cost minimization, as well the effect of the depot location (see for example Daganzo, 1984c; Newell and Daganzo, 1986; Daganzo, 1987; Langevin and Soumis, 1989; Robusté et al., 1990; Hall, 1996). Larson and Odoni (1981) provide useful insights for the multiroute problem, whereas a generalization of the TSP formula for zones of different shapes is provided in Daganzo (1984a).

Some authors (for example Adebisi and Hurdle, 1982; Aldaihani et al., 2004; Quadrifoglio, Hall, and Dessouky, 2005) adapt a model for fixed line bus systems to flexible services (i.e., services in which the buses can deviate from their predefined path to serve requests off the route). In those cases the decision variable usually considered is the headway between two successive vehicles or the slack in the schedule. This kind of service is different from our DRT system (see section 3),

since in our case there are no predefined paths and so headways cannot be defined. For this reason, a model for conventional transit system cannot be used in our case.

In the 1970's due to the diffusion of paratransit services, some researchers proposed different methodologies to model simplified variants of a DRT system in order to compare them with conventional bus line networks (Ward, 1975). The issue of the design of an integrated urban public transportation system was investigated by Batchelder and Kullman (1977). However in this case the model for the dial-a-ride system was based on computer simulations calibrated on real datasets. Wilson and Hendrickson (1980) focus on performance models, where the decision variable is related to the quality of the service, and provide an excellent comparative analysis on the different methodologies that have been proposed. They also report from previous unpublished research empirical models for the determination of the number of vehicles that were calibrated on real data. It is well known that empirical models are difficult to use in a context that is different from the one upon which they have been calibrated.

Another modeled system is “many-to-one”, where there are many origins and a single destination. In this case the service area is usually divided in several zones and in each of those only one vehicle can operate. Each vehicle collects the request in its zone and delivers them to a central location. It is straightforward to see that in this case it is possible to decompose the problem into several smaller TSP, and to successfully apply the previously mentioned “cluster-first, route-second” methodologies for the estimation of both the number of vehicles and the distance travelled. The decision variable in this case can be the size of the zone or the capacity of the vehicle. One of the first applications is reported in Daganzo et al. (1977). Chang and Schonfeld (1991) and Chang and Lee (1993) propose an equilibrium model for a many-to-one service that takes into account the demand elasticity.

Some research has focused on “many-to-many” systems, where there many origins and destinations. Arrillaga and Medville (1973) and Flushberg and Wilson (1976) make use of regression models, whereas Lerman and Wilson (1974) and Daganzo (1978) propose stochastic models in which the customer's arrival at a stop is a Poisson distributed queuing process. The general drawback of the former approach is the limitation of the validity range depending on the

datasets used for the calibration, whereas the latter may present problems in the case of uncongested systems. Daganzo (1984b) performs a comparison of fixed and flexible transit systems by modeling their costs. In this work, door to door service is a limiting case of a jitney service (i.e., the considered routing strategy consists of dispatching vehicles with constant headways and the stops without waiting passengers are skipped). An interesting theoretical discussion is provided in Stein (1978a, 1978b) where on the basis of a probabilistic analysis a class of scheduling rules is suggested. The outcome is that a decomposition algorithm in which buses serve a small zone and passengers across different zones have to transfer seems to asymptotically outperform systems in which the vehicles can travel in the whole area and a customer is inserted on the basis of the cost minimization.

Our goal is to model a many-to-many demand responsive transit service without predefined itineraries and schedules. In this case, the fleet has to be dispatched exclusively on the basis of the list of requests, like in taxicab systems, the difference being the possibility of serving customers with some detours in order to share the ride. We believe that this kind of service is of particular interest for the possibility of offering a high quality service with an efficient allocation of the resources. To achieve this, we will model a service in which time windows are associated with each pickup and delivery point.

Our definition of time window is different from the notion of “time deadline” that can be found in previous works, for example concerning hauling services (Hall, 1996). Although Daganzo (1987) modelled a distribution problem considering time windows associated with each delivery point, the suggested methodology (place each request in a time bin and then proceed with a “cluster-first, route-second” approach) is not suitable when temporal constraints are tight as in the case we are considering. We need a procedure that is not easily derivable from existing methodologies. For example, comparing our problem to the previously discussed ones, we can see that in our case it is impossible to model it as a fixed-line service since we cannot define a “path” or a “headway” between the vehicles. On the other hand, the joint need of avoiding transfers for any pair of pickup and delivery points and of limiting the maximum ride time for every customer prevents us from dividing the area into several service zones served by a single vehicle, so that also a “cluster-first, route-second” model is not appropriate.

3. SPECIFICATION OF THE STUDIED SYSTEM

In the following we will start by partially adopting the operating scenario described by Jaw et al. (1986). Our demand responsive transit (DRT) system consists of a fleet of vehicles with no predefined schedules. The vehicles travel at a constant speed and cannot idle. We later show where relaxing the no idling assumption and considering a more idealized scenario simplifies the problem. The service time at the locations is zero and we do not consider capacity constraints since in most 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. We assume that the coordinates of the pickup and the delivery points are random variables drawn from the same distribution. Hence, given this distribution, it is possible to compute the distribution of the Euclidean travel distances between any pair of points. Let $L(A,B)$ be a random variable from the latter distribution, representing the distance between point A and point B . In order to ensure an acceptable quality of the service, the vehicle has to pick up the customer no earlier than the pickup time and no later than a specified time interval from the pickup time. The vehicles cannot pick up a customer earlier than the pickup time because customers may not be there at that time. Also the maximum length of the trip must be somewhat limited. To do this, we fix a maximum wait state WS , which is the same for all the customers, and we compute a maximum ride time MRT_k for each request k . The maximum ride time is defined as an increasing function of the direct ride time (i.e., the time needed to serve the request without deviations). Since the vehicles travel at a constant speed v , the direct ride time is simply

$\frac{L(D_k, P_k)}{v}$ assuming the Euclidean metric, where P_k and D_k are the pickup and the delivery points

of request k . We can now compute the maximum ride time of each customer in the following way, where a and b are two parameters that are specified by the scheduler, with $a \geq 0$ and $b \geq 1$:

$$MRT_k = a + b \cdot \frac{L(P_k, D_k)}{v}$$

The above scheduling constraints related to the maximum wait state and maximum ride time for each request k define the quality of the service. The most practical way to take them into account in the scheduling process is to define time windows for all the pickup and delivery locations. Let EPT_k be the earliest pickup time requested by customer k . Then, let (EPT_k, LPT_k) and (EDT_k, LDT_k) be the time windows associated with the pickup and delivery times for customer k , respectively. It is possible to define these time windows on the basis of $\frac{L(D_k, P_k)}{v}$, WS and MRT_k in several different ways, each method having benefits and drawbacks that are discussed in Diana and Dessouky (2004). In this paper we use the following method to compute the time windows (see fig. 1).

$$LPT_k = EPT_k + WS$$

$$EDT_k = EPT_k + \frac{L(P_k, D_k)}{v}$$

$$LDT_k = EPT_k + MRT_k = EPT_k + a + b \frac{L(P_k, D_k)}{v}$$

Fig. 1.

Jaw et al. (1986) also consider the possibility of letting the customer specify the delivery time. In this case, the pickup time and the time windows are defined in a very similar way. The generalization of our methodology in this sense is straightforward.

4. A MODEL FOR ESTIMATING THE REQUIRED NUMBER OF VEHICLES

4.1. The expected number of vehicles

We have a list of n requests scattered in a service area. Our objective is to estimate the number of vehicles needed to serve these requests using the DRT system introduced in the previous section.

Let r_m be the probability of serving a set of m requests out of the n total requests with the same vehicle. By the above definition of the time windows, $r_1 = 1$. That is, each request can be satisfied if assigned to a vehicle. If for example we state that each vehicle cannot serve more than two requests, then there will be on average $\frac{n}{2}r_2$ vehicles that serve two requests and $n \cdot (1 - r_2)$ that serve the remainder. The expected total number of vehicles $E(z)$ needed to serve n requests is then

$$E(z) = \frac{n}{2}r_2 + n \cdot (1 - r_2)$$

If now we suppose that each vehicle can serve three requests, there will be on average $\frac{n}{3}r_3$ vehicles that serve three requests, $\frac{n}{2} \cdot r_2(1 - r_3)$ that serve two requests (where $r_2(1 - r_3)$ is the joint probability of serving two requests with a vehicle that could not serve three of them) and finally $n \cdot (1 - r_2)(1 - r_3)$ that serve only one request. Thus, the expected number of vehicles is

$$E(z) = \frac{n}{3}r_3 + \frac{n}{2}r_2(1 - r_3) + n \cdot (1 - r_2)(1 - r_3)$$

The expected number of vehicles needed to serve n requests can be computed generalizing the above equation:

$$E(z) = n \sum_{i=1}^n \frac{r_i}{i} \prod_{j=i+1}^n (1 - r_j) \quad [1]$$

It can be noted that the succession of the probabilities r_1, r_2, \dots, r_n rapidly converges to zero so that we need to determine only the first m values, with $m \ll n$. In the next section we will discuss how we can estimate the values of r_2, \dots, r_m .

4.2. The probability of serving m requests with one vehicle

4.2.1. The general case

From the definition of our problem, if one vehicle has to serve m requests it will have to visit $2m$ nodes (m pickups and m deliveries). Theoretically speaking there are $(2m)!$ possible visiting sequences, and we should compute the probability associated to each one. If we assume that the fleet dispatching process seeks for cost minimization, then the scheduler would choose the visiting sequence that maximizes the possibility of serving all the m requests. It follows that r_m would simply be the maximum of all the probabilities of success that are associated to the $(2m)!$ possible visiting sequences. However the presence of the pairing constraints (each pickup point must be visited before the corresponding delivery point) limits the number of feasible sequences (that is, of the sequences that have probability greater than zero) to $\frac{(2m)!}{2^m}$.

Let us focus our attention on the easiest case, that is for $m = 2$. We want to compute the probability of success in serving with one vehicle any pair of requests (say, 1 and 2) among the n requests waiting to be served. The vehicle must then visit four nodes: the pickup and delivery point of the first and of the second request each one having the above defined time window. We will indicate these points with P_1 , D_1 , P_2 and D_2 respectively. Considering the pairing constraint, the feasible sequences are only the following six:

$$P_1 D_1 P_2 D_2 \quad P_1 P_2 D_1 D_2 \quad P_1 P_2 D_2 D_1 \quad P_2 D_2 P_1 D_1 \quad P_2 P_1 D_2 D_1 \quad P_2 P_1 D_1 D_2$$

Now, we assume that r_2 is equal to the probability of realizing the most likely sequence among the above six. Each sequence is determined by three different events: for example, the first one is feasible if and only if we can serve first P_1 and then D_1 , D_1 and then P_2 , and P_2 and then D_2 . Since in section 3 we assumed that the location of any point is not related to the location of all the others, the travel times of these three events are independent. However, the arrival time to P_2 is dependent on the travel time of the first two legs. In order to simplify the computation of the joint

probability of the realization of the above sequence (i.e., P_1 to D_1 to P_2 to D_2) we assume that it is the product of the probabilities of the single events. This assumption of independence of the events related to a sequence overlooks the links between the arrival time at a node and the departure time from the same node. It is likely to be a more severe limitation as the time window width decreases and the vehicle is running late.

We will refer to the probabilities of the single events in a sequence using the term “elementary probabilities” and pd_{ij} indicates the probability of success in visiting the pickup point (p) of request i and then the delivery point (d) of request j . We will have similarly dp_{ij} , pp_{ij} and dd_{ij} . We can now express r_2 as a function of those quantities:

$$r_2 = \max \left(pd_{11} \cdot dp_{12} \cdot pd_{22}, pp_{12} \cdot pd_{21} \cdot dd_{12}, pp_{12} \cdot pd_{22} \cdot dd_{21}, \right. \\ \left. pd_{22} \cdot dp_{21} \cdot pd_{11}, pp_{21} \cdot pd_{12} \cdot dd_{21}, pp_{21} \cdot pd_{11} \cdot dd_{12} \right) \quad [2]$$

Considering again the definition of the time windows, we can say that $pd_{ii} = 1$ for every i . To determine all the other elementary probabilities we can proceed as follows. As an example, we will extensively show the procedure of computing dp_{ij} and we will give only the results for the other three cases since the steps are very similar.

Since the nodes D_i and P_j have a time window, the vehicle can serve both only if the travel time $\frac{L(D_i, P_j)}{v}$ between them is within a certain range. The random variable $L(D_i, P_j)$ and the constant v have been introduced in section 3. The upper limit of this range is reached if the vehicle visits D_i at the earliest time and P_j at the latest. If the vehicle is not allowed to idle, there is also a lower limit represented by the trip duration when the vehicle visits D_i at the latest time and P_j at the earliest. Figure 2 shows the relationship between these two limits and the time windows. The following equation translates this graphical relationship into a mathematical expression:

$$EPT_j - \left(EPT_i + a + b \cdot \frac{L(P_i, D_i)}{v} \right) \leq \frac{L(D_i, P_j)}{v} \leq EPT_j + WS - \left(EPT_i + \frac{L(P_i, D_i)}{v} \right)$$

This interval can be rewritten as

$$\begin{aligned} v \cdot (EPT_j - EPT_i - a) &\leq L(D_i, P_j) + bL(P_i, D_i) \\ L(D_i, P_j) + L(P_i, D_i) &\leq v \cdot (EPT_j - EPT_i + WS) \end{aligned} \quad [3]$$

Fig. 2.

For the other three elementary probabilities, the procedure is the same, and it is sufficient to change the time windows to be considered. The probability intervals associated to the other three elementary probabilities will then be

$$pp_{ij} : v \cdot (EPT_j - EPT_i - WS) \leq L(P_i, P_j) \leq v \cdot (EPT_j - EPT_i + WS) \quad [4]$$

$$\begin{aligned} pd_{ij} : v \cdot (EPT_j - EPT_i - WS) &\leq L(P_i, D_j) - L(P_i, D_i) \\ L(P_i, D_j) - bL(P_i, D_i) &\leq v \cdot (EPT_j - EPT_i + a) \end{aligned} \quad [5]$$

$$\begin{aligned} dd_{ij} : v \cdot (EPT_j - EPT_i - a) &\leq L(D_i, D_j) + bL(P_i, D_i) - L(P_j, D_j) \\ L(D_i, P_j) + L(P_i, D_i) - bL(P_j, D_j) &\leq v \cdot (EPT_j - EPT_i + a) \end{aligned} \quad [6]$$

In the above equations, EPT_i , EPT_j and $L(\bullet, \bullet)$ are random variables, whereas v , a , b and WS are constants.

4.2.2. Problem reduction steps

On the basis of equations [3] to [6], given the distributions of the random variables, it is theoretically possible to define the associated probability intervals, and thus the values of the elementary probabilities. However, in order to have a computationally tractable problem, it is necessary to make some simplifying assumptions concerning these distributions. In the following we will assume that EPT_i and EPT_j are drawn from the same distribution, as well as $L(\bullet, \bullet)$ for every value of the argument.

Another issue concerns the number of times we need to apply this procedure to compute all the elementary probabilities and the number of sequences to be considered when m is increasing. From equation [2] we see that we need to repeat eight times the above procedure for computing these eight elementary probabilities: $dp_{12}, pp_{12}, pd_{21}, dd_{12}, dp_{21}, pp_{21}, pd_{12}$ and dd_{21} . Furthermore, as m increases the number of elementary probabilities that need to be computed explodes.

In appendix 1 we show in detail the approach we propose to reduce the problem of computing the elementary probabilities. This consists of two main steps: the redefinition of equations [3] to [6] in order to be able to solve them and the reduction of the number of elementary probabilities considered. Each of these steps is presented in a separate section of the appendix. More restrictive assumptions are needed in order to achieve both these results, and their impact on the model applicability is also briefly discussed. The third section of appendix 1 presents the derivation of a closed-form equation for computing the probabilities r_m . However, this third step in the problem reduction process is valid only for a particular case so it is not considered in the computational experiments we present in the next section.

5. COMPUTATIONAL EXPERIMENTS

5.1. Distributions of the time intervals between pickup times

In order to implement our model, we need to specify the probability density function of the time intervals between two successive pickup times, $f(g)$. We assume Poisson arrivals. Hence, it follows that $f(g)$ is an exponential distribution with parameter $\lambda = 1/E(g)$. Let the distribution $f(hg)$ be the time interval between h successive pickup times. Hence, the distribution $f(hg)$ that appears in equations [10]-[13] of Appendix 1 becomes a gamma distribution with parameters $E(g)$ and h .

5.2. Distribution of the leg lengths in a vehicle route

We also need to specify the probability density function of the distance between two successive points in a route served by one vehicle, $f(L(\bullet, \bullet))$. In the following we will assume a complete

randomness for the spatial point pattern. This implies that the number of service points in any planar region with area A follows a Poisson distribution with parameter $\lambda = N/A$, N being the number of service points, and that the point coordinates are an independent random sample from a uniform distribution.

Even under this assumption, it is not straightforward to represent the probability density function of the distance between two successive points of a vehicle route. Appendix 2 reports some approximations that are based on previous research results and discusses why they are ill-suited to our case. The contents of this section and of appendix 2 are the outcome of some trial-and-error experiments.

In the following we present the procedure we used to find an approximation for $f(L(\bullet, \bullet))$, that builds on the results presented in appendix 2. In fact, empirical evidence based on the simulations suggested a lognormal distribution for $f(L(\bullet, \bullet))$. In the next section we show some sample plots of $f(L(\bullet, \bullet))$. We next describe our approximation for $E(L(\bullet, \bullet))$ and $VAR(L(\bullet, \bullet))$.

Let us start by illustrating the procedure for computing $E(L(\bullet, \bullet))$. Considering the case limit in which there are no time windows, if we do not take into account the precedence constraints our problem is reduced to a standard Traveling Salesman Problem (TSP). Previous research showed that when the number of points p is large the length \bar{L}_T of a TSP tour, assuming Euclidean metric and a square area A , is

$$\bar{L}_T \approx 0.75 \cdot \sqrt{A} \sqrt{p}$$

Assuming that in our problem the vehicles are routed like in a TSP, if each vehicle serves m requests then it has to visit m pickup and m delivery locations starting and coming back to the depot. So, the expected length of the tour would be

$$\bar{L}_T \approx 0.75 \cdot \sqrt{A} \sqrt{2m+1}$$

and the mean length \bar{L} of each leg, i.e., the mean distance between any two points in the route, is

$$\bar{L} \approx \frac{0.75 \cdot \sqrt{A} \sqrt{2m+1}}{2m+1} = \frac{0.75 \cdot \sqrt{A}}{\sqrt{2m+1}} \quad [7]$$

Equation [7] underestimates $E(L(\bullet, \bullet))$ since it does not take into account the effect of the pairing constraint, and for this we modify it as follows. Let us define the operator $\rho(r_m)$, that for a given r_m indicates how many times the elementary probability pd_{xx} is contained in the most likely sequence that is used to compute r_m itself. For example, considering equation [2], if the most likely sequence is the first of the six listed then $\rho(r_2) = 2$; if it is the second then $\rho(r_2) = 0$, if it is the last then $\rho(r_2) = 1$ and so on. The tour of the vehicle reaching $2m+2$ nodes is composed of $2m+1$ legs. Of these, we can say that $\rho(r_m)$ are traveled to serve directly a request. Now we assume that for these legs the mean length is only influenced by the pairing constraint. Since the pickup and delivery points are drawn from the same distribution, this mean length is simply the value shown in the second row of Table 3 in Appendix 2. For the remaining $2m+1-\rho(r_m)$ legs, we overlook the effect of the pairing constraint, so that the corresponding mean length is given by equation [7]. Hence, the mean length of a leg of the tour is the weighted average of two different estimates for $E(L(\bullet, \bullet))$:

$$E(L(\bullet, \bullet)) = \frac{\rho(r_m)}{2m+1} 0.52 \cdot \sqrt{A} + \left(1 - \frac{\rho(r_m)}{2m+1}\right) \frac{0.75 \cdot \sqrt{A} \sqrt{2m+2-\rho(r_m)}}{2m+1-\rho(r_m)} \quad [8]$$

In formulating the above equation we assumed that it is possible to take into account the effect of the pairing constraint only on a subset of the legs, whereas $E(L(\bullet, \bullet))$ of the remainder is only influenced by the TSP-like routing. It is likely that this distinction is more blurred in reality, so that the expected length of all the legs is influenced by both the pairing constraint and the TSP-like routing in different proportions, but we believe this is a second-order effect that can be overlooked as such.

In order to compute $VAR(L(\bullet, \bullet))$, we suppose that $f(L(\bullet, \bullet))$ has the same coefficient of variance CV of the distribution shown in Table 3 in Appendix 2. Keeping the same notation, we compute

$$CV = \frac{\sqrt{VAR(d)}}{E(d)} = \frac{\sqrt{0.0615}}{0.5214} = 0.4756$$

and then we have

$$VAR(L(\bullet, \bullet)) = (CV \cdot E(L(\bullet, \bullet)))^2$$

Thus, we analytically derived the probability density functions $f(L(\bullet, \bullet))$ and then $f(L(\bullet, \bullet) + L(\bullet, \bullet))$, $f(L(\bullet, \bullet) - L(\bullet, \bullet))$, and $f(L(\bullet, \bullet) + L(\bullet, \bullet) - L(\bullet, \bullet))$ were derived through convolution. Finally, polynomial approximations were used to solve equations [10]-[13] of Appendix 1 through numerical integration.

5.3. Experimental design and results

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. The aforementioned complete spatial randomness assumption implies that the pickup and delivery points are independently and uniformly distributed over the square area. In both cases we used the above specified distributions of g and $L(\bullet, \bullet)$ and we varied the number of requests from 12 to 120 (corresponding to a mean value of g ranging from 10 to 1 minute).

We also considered different time windows. Since in our DRT system the time window width directly affects the quality of the service, this sensitivity analysis is the key to assess the trade off between a higher quality of service and the corresponding increase of the costs, in terms of a greater number of vehicles needed. In order to simplify the presentation of our results and their subsequent analysis, in the following we will allow the vehicle to idle but we will keep $b = 1$. We show in Appendix 1 that we could as well consider a system in which either $b > 1$ or the vehicles are not allowed to idle. Furthermore, we set $a = WS$. These assumptions imply that the pickup and the delivery time windows are the same for all the requests.

The lognormal model we introduced in section 5.2 seems to satisfactorily approximate $f(L(\bullet, \bullet))$ in those cases. We report in Figure 3 the plots of the sampled distribution of the leg lengths when we have 120 requests and time windows of 10 and of 30 minutes, together with their respective approximation when $m = 7$.

Fig. 3

As mentioned in section 4.1, when we compute the probabilities of serving 2, 3, ..., m requests with one vehicle, we do not need to fix $m = n$. In fact, the values for r_i are decreasing when i increases, and the corresponding addends in equation [1] become less and less influential on the value of $E(z)$. Thus, we compute the series of probabilities r_2, r_3, \dots, r_m until the expected number of vehicles cannot be changed by the addends related to the probabilities r_{m+1}, \dots, r_n . However we also tested a more stringent stopping criterion. Since the scheduling horizon is of two hours, we impose that the time needed to serve m requests with a vehicle must be less than the deadline of delivering the latest request. Hence, we compute the probabilities only until m satisfies the following inequality:

$$\frac{2m}{v} E(L(\bullet, \bullet)) < m \cdot E(hg) + TW + \frac{1}{v} E(L(\bullet, \bullet))$$

where $E(hg)$ is the expectation of the $f(hg)$ distribution, v is the speed of the vehicles and TW is the time window length.

When computing r_2, \dots, r_m , we noticed that the most likely sequence is always the following one: $P_1 D_1 P_2 D_2 \dots P_m D_m$. This consistently occurred throughout the entire experimental plan. However, it depends on the value of the elementary probabilities and cannot be shown to be a general rule. Nevertheless, when solving these problems we could always take the above sequence as the most likely. This leads to a drastic computational simplification since equation [14] of Appendix 1 and its extensions that we use to compute the elementary probabilities are reduced to the following form:

$$r_i = pd_0^i \cdot dp_{n/i}^{i-1} = dp_{n/i}^{i-1}$$

Also the mean leg length formula [8] is simplified following this approach since we have $\rho(r_i) = i$ for every $i = 1, \dots, m$. Hence, equation [8] becomes

$$E(L(\bullet, \bullet)) = \frac{m}{2m+1} \cdot 0.52 \cdot \sqrt{A} + \left(1 - \frac{m}{2m+1}\right) \frac{0.75 \cdot \sqrt{A} \sqrt{m+2}}{m+1}$$

The results of the computational experiments are shown in Table 1. In order to benchmark our planning model, we compare it to a simulation approach that requires determining the complete daily schedule. In the simulation model, requests were generated that followed the above mentioned distributions. The requests were scheduled using a parallel regret insertion algorithm (Diana and Dessouky, 2004). The regret insertion method allowed us to find the minimum number of vehicles required to service all the requests. In order to do this, we performed the first run of the algorithm with a very high number of vehicles, and we progressively lowered this number in the successive runs until some requests could not be scheduled.

Table 1

In the table it can be seen that the gap between the results from the model and those from the simulation, shown in brackets, is almost always less than two vehicles. Only when we have to serve 120 requests and the time windows are of 30 minutes do we overestimate the number of needed vehicles by over 2.1 in comparison with the simulation results. We believe that this is due to the approximation of the leg lengths that we used. We report in Table 2 the values of $E(L(\bullet, \bullet))$ computed from equation [8] and those derived from the schedules of the simulation; it can be seen that we overestimate the expected value of the leg lengths with larger time windows when we have to serve a higher number of requests. In particular, when we have 120 requests and the time windows are of 30 minutes we estimate a mean leg length of 3.72 miles, whereas the one from the simulation is 2.70 miles. The model overestimates the mean leg length in this case because there is a significant amount of ridesharing when there are a large number of requests and a wide time window. With a significant amount of ridesharing our approach in Equation [8]

under weights the trips that follow the TSP tour versus those that have a longer expected length of $0.52 \cdot 10 = 5.2$ miles. In this scenario, only a small number of trips go directly from pickup to delivery. Thus, only a small fraction of trips have a mean length of 5.2 miles. This suggests deriving a new weighting scheme in Equation [8] when there is a significant amount of ridesharing.

Table 2

Considering that a better approximation of $f(L(\bullet, \bullet))$ would even improve the model results, we believe that the proposed methodology is an effective way to quickly estimate the number of vehicles needed to provide a DRT under a fairly broad range of cases (systems with different levels of demand and different quality requirements).

6. CONCLUSIONS

In this paper we presented a continuous approximation model to forecast the number of vehicles needed to operate a demand responsive transit service. In contrast with current mathematical programming techniques, our approach simply requires the knowledge of the demand density over the service area since it may be hard or even impossible to have more detailed data in the planning phase. Computational results showed that the proposed methodology can provide reliable results under different circumstances. A critical point is the approximation of the distribution of the leg lengths that can alter the results when the time windows are wide, and more research is needed in this point in order to improve the performances of the proposed model.

The interest in using an approximation model lies in the possibility for the planner of performing sensitivity analyses through the construction of several different scenarios. In this way, the choice of the best compromise between quality of service and financial resources is much more effective. We believe that an application of particular interest of this model is the study of the tradeoff between the number of vehicles needed and the time windows associated with the locations, in analogy with what is shown in Table 1. This is a research field that deserves more

attention and that may be a key issue in developing DRT services that are more cost-effective but still satisfying for the customers.

The problem that we studied in this paper, as described in section 3, is sufficiently general to envision the application of our methodology in different contexts, for example in problems of distribution of goods in which there are severe time constraints. Another useful generalization of the present work might be the inclusion of the proposed methodology in a demand-supply equilibrium model for a general DRT system, similarly to what has been proposed by Chang and Schonfeld (1991) and Chang and Lee (1993) for the specific case of a deviation service.

ACKNOWLEDGEMENTS

This research was made possible by a grant from the Turin Transit Operator GTT (*Gruppo Torinese Trasporti*), Italy. We also acknowledge PATH (California Partners for Advanced Transit and Highways) and CALTRANS (California Department of Transportation) for their support.

REFERENCES

- Adebisi, O. and Hurdle, V.F. (1982) Comparing fixed-route and flexible-route strategies for intraurban bus transit. *Transportation Research Record 854*, Transportation Research Board, National Research Council, Washington D.C., 37-43.
- Aldaihani, M.M., Quadrifoglio, L., Dessouky, M. and Hall, R.W. (2004) Network design for a grid hybrid transit service. *Transportation Research*, 38A, 511-530.
- Arrillaga, B. and Medville, D.M. (1974) Demand, supply and cost modelling framework for demand-responsive transportation systems. *Transportation Research Board Special Report 147*, 32-47.
- Batchelder, J.H. and Kullman, B.C. (1977) Analysis of integrated urban public transportation systems. *Transportation Research Record 639*, Transportation Research Board, National Research Council, Washington D.C., 25-29.

- 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.
- Chang, S.K.J. and Lee, C.J. (1993) Welfare Comparison of Fixed- and Flexible-Route Bus Systems. *Transportation Research Record 1390*, Transportation Research Board, National Research Council, Washington D.C., 16-22.
- Chang, S.K. and Schonfeld, P.M. (1991) Optimization models for comparing conventional and subscription bus feeder services. *Transportation Science*, 25, 281-298.
- Christofides, N. and Eilon, S. (1969) Expected distances in distribution problems. *Operational Research Quarterly*, 20, 437-443.
- Daganzo, C.F. (1978) An approximate analytic model of many-to-many demand responsive transportation systems. *Transportation Research*, 12, 325-333.
- Daganzo, C.F. (1984a) The length of tours in zones of different shapes. *Transportation Research B*, 18B, 135-145.
- Daganzo, C.F. (1984b) Checkpoint dial-a-ride systems. *Transportation Research B*, 18B, 318-327.
- Daganzo, C.F. (1984c) The distance traveled to visit N points with a maximum of C stops per vehicle: an analytic model and an application. *Transportation Science*, 18, 331-350.
- Daganzo, C.F., (1987) Modeling distribution problems with time windows: Part I. *Transportation Science*, 21, 171-179.
- Daganzo, C.F. (1991) *Logistic Systems Analysis*. Lecture Notes in Economics and Mathematical Systems 361, Springer-Verlag, Heidelberg.
- Daganzo, C.F., Hendrickson, C.T. and Wilson, N.H.M. (1977) An approximate analytic model of many-to-one demand responsive transportation systems. *Proceedings of the 7th international symposium on transportation and traffic theory* eds T. Sasaki and T. Yamaoka, pp. 743-772, Kyoto.
- Del Castillo, J.M. (1999) A heuristic for the traveling salesman problem based on a continuous approximation. *Transportation Research B*, 33B, 123-152.
- Desaulniers, G., Desrosiers, J., Erdmann, A., Solomon, M.M. and Soumis, F. (2001) The VRP with pickup and delivery. *The Vehicle Routing Problem* eds P. Toth and D. Vigo, pp. 225-242, SIAM Monographs on Discrete Mathematics and Applications 9, Philadelphia.

- Diana, M. and Dessouky, M.M. (2004) A new regret insertion heuristic for solving large-scale dial-a-ride problems with time windows. *Transportation Research B*, 38B, 539-557.
- Donnelly, K.P. (1978) Simulations to determine the variance and edge effect of total nearest-neighbour distance. *Simulation Studies in Archaeology* ed I. Hodder, pp. 91-95, Cambridge University Press, Cambridge.
- Eilon, S., Watson-Gandy, C. and Christofides, N. (1971) Expected distances in distribution problems. *Distribution Management: Mathematical Modelling and Practical Analysis*, pp. 151-179, Griffin, London.
- Flusberg, M. and Wilson, N.H.M. (1976) A descriptive supply model for demand responsive transportation system planning. *Proceedings of the 17th Annual Meeting of the Transportation Research Forum*, 425-431.
- Hall, R.W. (1996) Pickup and delivery systems for overnight carriers. *Transportation Research A*, 30A, 173-187.
- 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.
- Langevin, A., Mbaraga, P. and Campbell, J.F. (1996) Continuous approximation models in freight distribution: an overview. *Transportation Research B*, 30B, 163-188.
- Langevin A. and Soumis, F. (1989) Design of multiple-vehicle delivery tours satisfying time constraints. *Transportation Research B*, 23B, 123-138.
- Larson, R.C. and Odoni, A.R. (1981) *Urban Operations Research*. Prentice-Hall, Englewood Cliffs.
- Lave, R.E., Teal, R. and Piras, P. (1996) *A handbook for acquiring demand-responsive transit software*. Transit Cooperative Research Program Report #18, Transportation Research Board, Washington, D.C.
- Lerman, S.R. and Wilson, N.H.M. (1974) Analytic model for predicting dial-a-ride system performance. *Transportation Research Board Special Report 147*, 48-53.
- Mathai, A.M. (1999) The nearest neighbor problem on a plane from Poisson arrivals of random points. *An Introduction to Geometrical Probability*, pp. 223-228, Gordon and Breach, Amsterdam.

- Newell, G.F. and Daganzo, C.F. (1986) Design of multiple-vehicle delivery tours – I. A ring radial network. *Transportation Research B*, 20B, 345-363.
- Palmer, K, Dessouky, M. M., and Abdelmaguid, T. (2004) Impacts of management practices and advanced technologies on demand responsive transit systems. *Transportation Research, Part A: Policy and Practice*, 38, 495-509.
- Quadrifoglio, L., Dessouky, M. M. and Ordóñez, F. (2005) A simulation study of demand responsive transit system design. submitted for publication.
- Quadrifoglio, L, Hall, R. W., and Dessouky, M. M. (2005) Design of Mobility Allowance Shuttle Transit (MAST) Services: bounds on the maximum longitudinal velocity, under revision *Transportation Science*.
- Robusté, F., Daganzo, C.F. and Souleyrette, R.R. (1990) Implementing vehicle routing models. *Transportation Research B*, 24B, 263-286.
- Savelsbergh, M.W.P. and Sol, M. (1995) The general pickup and delivery problem. *Transportation Science*, 29, 17-29.
- Solomon, M.M. (1987) Algorithms for the vehicle routing and scheduling problems with time windows constraints. *Operations Research*, 35, 254-265.
- Stein, D.M. (1978a) Scheduling dial-a-ride transportation problems. *Transportation Science*, 12, 232-249.
- Stein, D.M. (1978b) An asymptotic probabilistic analysis of a routing problem. *Mathematics of Operations Research*, 3, 89-101.
- Ward, D.E. (1975) *Theoretical comparison of fixed route bus and flexible route subscription bus service in low density areas*. Transportation Systems Center, USDOT.
- Wilson, N.H.M. and Hendrickson, C. (1980) Performance models of flexibly routed transportation services. *Transportation Research B*, 14B, 67-78.

APPENDIX 1. COMPUTATION OF THE ELEMENTARY PROBABILITIES

In this appendix we show the computation of the elementary probabilities under different assumptions.

A1.1. Distribution of the intervals between adjacent pickup times

Let us define a ranking order for the list of requests by ascending EPT . Define a random variable g that represents the temporal gap between the two earliest pickup times of the requests k and $k+1$, if k is the index of this ranking order:

$$g = EPT_{k+1} - EPT_k$$

In the general case, if two requests are at the k^{th} and at the $(k+h)^{th}$ place according this order, we can say that the temporal gap between their respective earliest pickup times can be represented by the distribution of the random variable $\sum_{k=1}^h g$. For simplicity, we will refer to this random

variable as hg . Since h is not a constant (i.e., the number of demand points is a random variable that follows a discrete distribution), we point out that the associated probability density function $f(hg)$ is not equal to $h \cdot f(g)$. It is now possible to redefine the elementary probabilities introduced in section 4.2 as a function of these interval gaps. Considering again the case $m = 2$, we have for example that for the first of the above listed six feasible sequences the vehicle should start visiting the pickup and the delivery node of the same request. The difference of order between the requests related to the two nodes is obviously 0 since both are from the same request. So in this case $h = 0$ and we can denote the corresponding probability as pd_0 . When leaving D_1 , the vehicle must arrive to P_2 . We assume that the time gap between the two requests is $(n/2) \cdot g$. We denote the corresponding elementary probability with $dp_{n/2}$, which is the probability of picking up the $(k+n/2)^{th}$ request after delivering the k^{th} . Defining in the same way all the remaining elementary probabilities, equation [2] can be rewritten as follows:

$$r_2 = \max \left(pd_0 \cdot dp_{n/2} \cdot pd_0, pp_{n/2} \cdot pd_{-n/2} \cdot dd_{n/2}, pp_{n/2} \cdot pd_0 \cdot dd_{-n/2}, \right. \\ \left. pd_0 \cdot dp_{-n/2} \cdot pd_0, pp_{-n/2} \cdot pd_{n/2} \cdot dd_{-n/2}, pp_{-n/2} \cdot pd_0 \cdot dd_{n/2} \right) \quad [9]$$

Nine different elementary probabilities appear in this formula. In the general case (that is, for m greater than 2), the interval gap is $c \cdot (n/m) \cdot g = hg$, where c is an integer constant comprised between $-(m-1)$ and $+(m-1)$. The number of elementary probabilities that must be determined is linearly increasing with m , being equal to $(8m-7)$. They are the following: $pp_{-(m-1) \cdot n/m}, \dots, pp_{-n/m},$

$pp_{n/m}, \dots, pp_{(m-1)\cdot n/m}, pd_{-(m-1)\cdot n/m}, \dots, pd_{-n/m}, pd_0, pd_{n/m}, \dots, pd_{(m-1)\cdot n/m}, dp_{-(m-1)\cdot n/m}, \dots, dp_{-n/m}, dp_{n/m}, \dots, dp_{(m-1)\cdot n/m}, dd_{-(m-1)\cdot n/m}, \dots, dd_{-n/m}, dd_{n/m}, \dots, dd_{(m-1)\cdot n/m}$. We can see that h assumes a value of zero only when there is a request that is served without deviations; in this case the corresponding elementary probability pd_0 is 1 by definition. In all the other cases (i.e. when $h \neq 0$), the definition of the elementary probabilities as a function of h allows us to compute more easily their values. In fact, assuming $b = 1$, equations [3] to [6] can be rewritten in this manner:

$$\begin{aligned} dp_h &: v \cdot (hg - a) \leq L(D_i, P_j) + L(P_i, D_i) \leq v \cdot (hg + WS) \\ pp_h &: v \cdot (hg - WS) \leq L(D_i, P_j) \leq v \cdot (hg + WS) \\ pd_h &: v \cdot (hg - WS) \leq L(D_i, P_j) - L(P_i, D_i) \leq v \cdot (hg + a) \\ dd_h &: v \cdot (hg - a) \leq L(D_i, P_j) + L(P_i, D_i) - L(P_j, D_j) \leq v \cdot (hg + a) \end{aligned}$$

Assuming $b = 1$ is useful to transform for each elementary probability the two inequalities in a probability interval. In order to have only one interval in the above equations, we could alternatively allow the vehicle to idle at every node. As we discussed in section 4.2, the lower bound would disappear in this case (if the travel time is too short, then the vehicle could wait at the second point until the time window is met) and the assumption of $b = 1$ could be relaxed.

It is reasonable to assume that the random variables $L(D_i, P_j)$ and (hg) are independent, as well $(L(D_i, P_j) + L(P_i, D_i))$ and (hg) , $(L(D_i, P_j) - L(P_i, D_i))$ and (hg) and $(L(D_i, P_j) + L(P_i, D_i) - L(P_j, D_j))$ and (hg) . Hence, it is possible to express the elementary probabilities as follows:

$$dp_h = \begin{cases} \int_0^\infty \int_{v(hg-a)}^{v(hg+WS)} f(L(D_i, P_j) + L(P_i, D_i)) \cdot f(hg) d(L(D_i, P_j) + L(P_i, D_i)) d(hg) & (h > 0) \\ \int_{-\infty}^0 \int_0^{v(hg+WS)} f(L(D_i, P_j) + L(P_i, D_i)) \cdot f(hg) d(L(D_i, P_j) + L(P_i, D_i)) d(hg) & (h < 0) \end{cases} \quad [10]$$

$$pp_h = \begin{cases} \int_0^\infty \int_{v(hg-WS)}^{v(hg+WS)} f(L(D_i, P_j)) \cdot f(hg) d(L(D_i, P_j)) d(hg) & (h > 0) \\ \int_{-\infty}^0 \int_0^{v(hg+WS)} f(L(D_i, P_j)) \cdot f(hg) d(L(D_i, P_j)) d(hg) & (h < 0) \end{cases} \quad [11]$$

$$pd_h = \begin{cases} \int_0^\infty \int_{v(hg-WS)}^{v(hg+a)} f(L(D_i, P_j) - L(P_i, D_i)) \cdot f(hg) d(L(D_i, P_j) - L(P_i, D_i)) d(hg) & (h > 0) \\ \int_{-\infty}^0 \int_{v(hg-WS)}^{v(hg+a)} f(L(D_i, P_j) - L(P_i, D_i)) \cdot f(hg) d(L(D_i, P_j) - L(P_i, D_i)) d(hg) & (h < 0) \end{cases} \quad [12]$$

$$dd_h = \begin{cases} \int_0^\infty \int_{v(hg-a)}^{v(hg+a)} f(L(D_i, P_j) + L(P_i, D_i) - L(P_j, D_j)) \cdot f(hg) \\ \quad d(L(D_i, P_j) + L(P_i, D_i) - L(P_j, D_j)) d(hg) & (h > 0) \\ \int_{-\infty}^0 \int_{v(hg-a)}^{v(hg+a)} f(L(D_i, P_j) + L(P_i, D_i) - L(P_j, D_j)) \cdot f(hg) \\ \quad d(L(D_i, P_j) + L(P_i, D_i) - L(P_j, D_j)) d(hg) & (h < 0) \end{cases} \quad [13]$$

In the preceding equation with $f(\bullet)$ we indicated the probability density functions of the random variables $L(D_i, P_j)$, $(L(D_i, P_j) + L(P_i, D_i))$, $(L(D_i, P_j) - L(P_i, D_i))$, $(L(D_i, P_j) + L(P_i, D_i) - L(P_j, D_j))$ and (hg) , that can be computed using convolution. Theoretically speaking, we could now compute the probabilities r_m . However, if m is increasing we still have the problem of the number of elements to consider in equation [9], that as aforesaid is equal to $\frac{(2m)!}{2^m}$ (in the general case there is one element for each feasible visiting sequence). The method to cope with this problem is shown in the next section.

A1.2. Serving the requests by ascending pickup time

Let us assume that the vehicle serves the requests following the ranking order we introduced in the previous section. That is if $EPT_i < EPT_j$ then node P_i will be visited before P_j . With reference to the above shown case for $m = 2$, we can say that the first three sequences will be feasible if request 2 follows request 1 in this ranking order, whereas the last three may occur only if request 1 follows request 2. Since these two events are mutually exclusive and have the same probability, we can then focus our attention on only one case assuming for example that request 1 precedes request 2. Thus we will only have to consider these three sequences:

$$P_1 D_1 P_2 D_2 \quad P_1 P_2 D_1 D_2 \quad P_1 P_2 D_2 D_1$$

and the probability of serving two requests with one vehicle becomes

$$r_2 = \max \left(pd_0^2 \cdot dp_{n/2}, pp_{n/2} \cdot pd_{-n/2} \cdot dd_{n/2}, pp_{n/2} \cdot pd_0 \cdot dd_{-n/2} \right) \quad [14]$$

For $m = 3$, we have 15 sequences to consider in order to compute r_3 :

$$r_3 = \max \left(pd_0^3 \cdot dp_{n/3}^2, pp_{n/3} \cdot pd_{-n/3} \cdot pd_0 \cdot dp_{n/3} \cdot dd_{n/3}, pp_{n/3} \cdot pd_0^2 \cdot dp_{n/3} \cdot dd_{-n/3}, \right. \\ pp_{n/3} \cdot pd_{-n/3}^2 \cdot dp_{2n/3} \cdot dd_{n/3}, pp_{n/3} \cdot pd_{-n/3} \cdot pd_0 \cdot dp_{2n/3} \cdot dd_{-n/3}, \\ pp_{n/3} \cdot pd_0^2 \cdot dp_{2n/3} \cdot dd_{-n/3}, pp_{n/3} \cdot pd_{-2n/3} \cdot pd_0 \cdot dp_{n/3} \cdot dd_{2n/3}, \\ pp_{n/3} \cdot pd_0^2 \cdot dp_{n/3} \cdot dd_{-2n/3}, pp_{n/3}^2 \cdot pd_{-2n/3} \cdot dd_{n/3}^2, pp_{n/3}^2 \cdot pd_{-2n/3} \cdot dd_{-n/3} \cdot dd_{2n/3}, \\ pp_{n/3}^2 \cdot pd_{-n/3} \cdot dd_{-n/3} \cdot dd_{2n/3}, pp_{n/3}^2 \cdot pd_{-n/3} \cdot dd_{-2n/3} \cdot dd_{n/3}, \\ \left. pp_{n/3}^2 \cdot pd_0 \cdot dd_{-2n/3} \cdot dd_{n/3}, pp_{n/3}^2 \cdot pd_0 \cdot dd_{-n/3}^2 \right)$$

In the general case, this approach allows us to dramatically reduce the number of sequences to be considered, now being $\frac{(2m)!}{2^m m!}$ instead of $\frac{(2m)!}{2^m}$. This number is still big when m is increasing.

However, it will be shown later that in most practical cases the maximum m to be considered is such that the computation of the corresponding r_m can be performed in a reasonable amount of computational time. Also the number of the elementary probabilities to be computed is reduced with this approach. It can be seen that for $m = 2$ we need to compute 5 of these, and $(4m-3)$ in the general case. They are $pp_{n/m}, pd_{-(m-1) \cdot n/m}, \dots, pd_{-n/m}, dp_{n/m}, \dots, dp_{(m-1) \cdot n/m}, dd_{-(m-1) \cdot n/m}, \dots, dd_{-n/m}, dd_{n/m}, \dots, dd_{(m-1) \cdot n/m}$. Furthermore, with a partially different analysis approach, there is the possibility of defining the elementary probabilities in such a way that the sequences need no more to be enumerated. In this case, it is possible to write a general equation in a compact form to compute r_m , instead of using an extension of equation [14]. In the following section we will discuss this simplified approach.

11.3. Considering only contemporary requests

Up to now, we tackled the problem of estimating the number of vehicles needed in a DRT system following a “global approach”, that is concurrently considering all the requests to be served within our analysis period. We believe that this is the method that can give the most accurate

results, but we have shown that this may lead to some complications when performing the calculations.

We can also study our problem using a “discrete approach” dividing our analysis period in small time intervals. In this case, all the requests whose *EPT* falls within a given interval can be considered having the same *EPT* that is the central value of the interval. For each cluster of requests it is easily possible to determine the number of vehicles that is needed following a simplified version of the above methodology. However it might be difficult to assess the number of vehicles needed over the entire planning horizon using a discrete approach so we will not consider this method in subsequent analyses. We present it here mainly for the purpose of giving a simple equation for r_m .

If we are considering only a subset of requests that can be considered having the same *EPT*, then it follows that the above defined g random variable is 0. The probability intervals associated to each elementary probability become the following:

$$dp : 0 \leq L(D_i, P_j) + L(P_i, D_i) \leq v \cdot WS$$

$$pp : 0 \leq L(D_i, P_j) \leq v \cdot WS$$

$$pd : -v \cdot WS \leq L(D_i, P_j) - L(P_i, D_i) \leq v \cdot a$$

$$dd : -v \cdot a \leq L(D_i, P_j) + L(P_i, D_i) - L(P_j, D_j) \leq v \cdot a$$

Since $g = 0$, the elementary probabilities are no more dependent on h and so we can drop the subscript. For any m we will have to consider only 4 elementary probabilities that can be computed, given the above probability intervals in the following manner:

$$dp = \int_0^{vWS} f(L(D_i, P_j) + L(P_i, D_i)) d(L(D_i, P_j) + L(P_i, D_i))$$

$$pp = \int_0^{vWS} f(L(D_i, P_j)) d(L(D_i, P_j))$$

$$pd = \int_{-vWS}^{va} f(L(D_i, P_j) - L(P_i, D_i)) d(L(D_i, P_j) - L(P_i, D_i))$$

$$dd = \int_{-va}^{va} f(L(D_i, P_j) + L(P_i, D_i) - L(P_j, D_j)) d(L(D_i, P_j) + L(P_i, D_i) - L(P_j, D_j))$$

Considering again the case $m = 2$, equation [5] can then be rewritten in the following way:

$$\begin{aligned} r_2 &= \max (pd \cdot dp \cdot pd, pp \cdot pd \cdot dd, pp \cdot pd \cdot dd) \\ &= pd \cdot \max (pd \cdot dp, pp \cdot dd) \end{aligned}$$

that for a generic value of m becomes

$$r_m = pd \cdot \max_{i=1..m} (pp^{i-1} \cdot pd^{m-i} \cdot dp^{m-i} \cdot dd^{i-1}).$$

Combining this equation with equation [1] gives a formula to compute the expected number of vehicles needed to serve n requests having the same pickup time, but different pickup and delivery locations.

APPENDIX 2. DISTRIBUTION OF THE DISTANCES OF RANDOM AND OF NEAREST-NEIGHBOR POINTS IN THE SERVICE AREA

In this appendix we discuss the viability of two possible approximations for the distribution of the leg lengths in a demand responsive service vehicle route. The first approach is to consider the distribution of the distance between any two random points in the service area $f(d)$ that can be obtained from the distributions of the coordinates of the points through convolution. Another possible strategy is to consider the (s -th) nearest-neighbor distance density $f(d_s)$ from a given point in the service area. In other words, either the distance d between any two points in the service area or between a point and its (s -th) nearest neighbor d_s could be used as an approximation of the distance between any two points in a vehicle route $L(\bullet, \bullet)$.

Under the complete spatial randomness hypothesis, some results related to the distributions $f(d)$ and $f(d_s)$ are available in the published literature. Christofides and Eilon (1969) and Eilon et al. (1971) derive the expected distance between two random points for different shapes of the service area. The distribution of the distances $f(d)$ and their mean $E(d)$ and variance $VAR(d)$ for the cases of points uniformly scattered over a unit service area are reported in Table 3. Spatial analysis textbooks such as Mathai (1999) report the probability density function $f(d_s)$ of the nearest, second-nearest, ..., s -th nearest point, assuming Poisson arrivals in a plane. However if we consider a finite area, there are boundary effects that alter these latter distributions. In fact, we would expect that the value of d_s is greater and is increasing when we consider points that are nearer the edge. Considering the case of the nearest neighbor ($s = 1$) and the related distribution of the distances d_1 , Donnelly (1978) determined correction terms through simulation for $E(d_1)$ and $VAR(d_1)$, that are sufficiently accurate when there are more than seven points and the shape of the region is sufficiently smooth. We report the expressions for $f(d_1)$, $E(d_1)$ and $VAR(d_1)$ both considering and not considering edge effects in Table 4.

Tables 3 and 4

In order to check the possibility of approximating $f(L(\bullet, \bullet))$ through either $f(d)$ or $f(d_1)$, we ran some simulations on standard problems (pickup and delivery points uniformly scattered in a unit square area). To schedule the vehicles in the simulation, we used a parallel regret insertion heuristic (Diana and Dessouky, 2004). The results showed that $E(L(\bullet, \bullet))$ is about 20% to 40% less than the value of $E(d)$ indicated in Table 3, and this gap is increasing when the time windows are larger and the density of the requests is higher. This is rather an intuitive result since relaxing the scheduling constraints leads to a more efficient routing of the vehicles. On the other hand, $E(d_1)$ seriously underestimates $E(L(\bullet, \bullet))$ when there are more than 3-4 requests to serve and the time window width is not too loose even if we include the correction terms proposed by Donnelly (1978). Also the shape of the sampled distribution is quite different from those reported in Tables 3 and 4.

To sum up, the considered approximations have proven to be rather poor for our purposes. When scheduling the service the vehicles are normally dispatched to the “best” point that satisfies the time, precedence, and coupling constraints in order to increase the efficiency of the system. The definition of “best” point obviously depends on the heuristic used to schedule the service, but in any case this is rather unlikely to be either a random or the nearest-neighbor point. In our case, the regret insertion algorithm tries to anticipate the insertion of requests that could be difficult to insert in a later stage of the process, as explained in detail in Diana and Dessouky (2004). One could argue that using a nearest-neighbor-based heuristic to schedule the service could allow for a better approximation of $f(L(\bullet, \bullet))$ through $f(d_i)$. However the inferiority of the nearest-neighbor rule over an insertion-based algorithm when we consider a routing problem with time windows is an established result (Solomon, 1987).

List of figures

Fig. 1. Time windows definition

Fig. 2. Computation of the elementary probability dp_{ij}

Fig. 3. Distribution of the leg lengths from the simulation and from the model when 120 requests are serviced and the time windows width TW is of 10 and 30 minutes

List of tables

Table 1. Number of required vehicles from the model and from the simulation (in brackets) when the demand density and the time windows width is changing

Table 2. Expected leg lengths from the model and from the simulation (in brackets) when the demand density and the time windows width is changing

Table 3. Distribution of the distances between any two points that are uniformly scattered

Table 4. Distribution of the nearest-neighbor distance from a given point in a unit square, assuming Poisson arrivals of N points

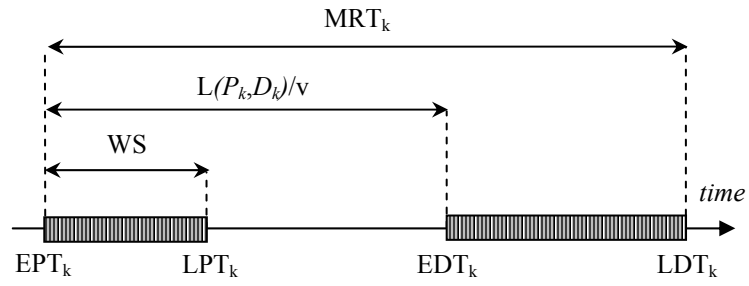


Fig. 1. Time windows definition

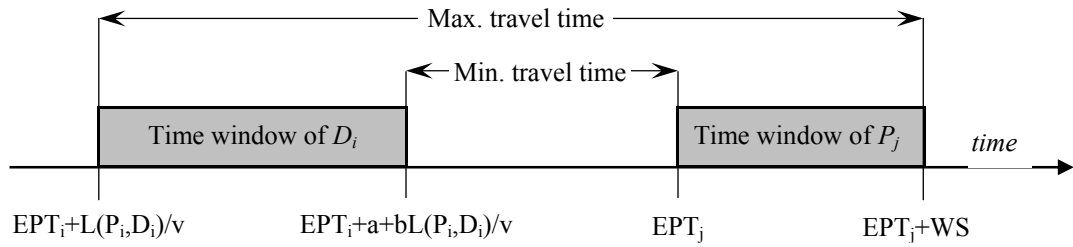


Fig. 2. Computation of the elementary probability dp_{ij}

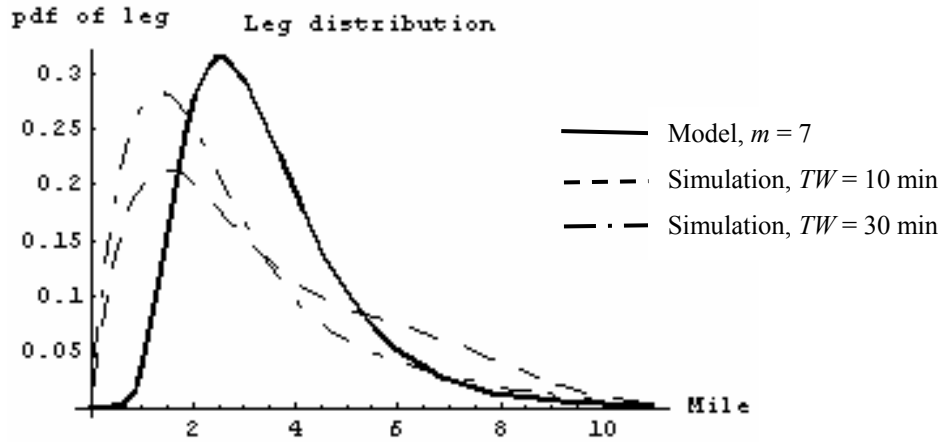


Fig. 3. Distribution of the leg lengths from the simulation and from the model when 120 requests are serviced and the time window width TW is 10 and 30 minutes

Table 1

Number of required vehicles from the model and from the simulation (in brackets) when the demand density and the time window width is changing

Time window	10 minutes	15 minutes	20 minutes	30 minutes
120 requests	22.9 (21.2)	18.7 (18.4)	17.6 (16.0)	17.2 (13.4)
60 requests	11.8 (13.0)	9.5 (11.0)	8.9 (9.6)	8.6 (8.0)
24 requests	5.2 (7.2)	4.1 (6.2)	3.7 (5.8)	3.4 (4.4)
12 requests	2.9 (4.4)	2.2 (3.2)	1.8 (3.2)	1.7 (2.6)

Table 2

Expected leg length in miles from the model and from the simulation (in brackets) when the demand density and the time window width is changing

Time window	10 minutes	15 minutes	20 minutes	30 minutes
120 requests	3.97 (3.51)	3.89 (3.27)	3.81 (3.03)	3.72 (2.70)
60 requests	4.08 (3.96)	3.99 (3.62)	3.90 (3.32)	3.81 (2.98)
24 requests	4.30 (4.14)	4.20 (4.12)	4.16 (3.93)	3.99 (3.80)
12 requests	4.45 (4.58)	4.22 (4.33)	4.22 (4.33)	4.09 (3.78)

Table 3

Distribution of the distances between any two points that are uniformly scattered in a unit square

$f(d)$	$\begin{cases} 2\pi d - 8d^2 + 2d^3 & : 0 \leq d \leq 1 \\ 4d \arcsin\left(\frac{2-d^2}{d^2}\right) + 8d\sqrt{d^2-1} - 4d - 2d^2 & : 1 < d \leq \sqrt{2} \\ 0 & : \text{otherwise} \end{cases}$
$E(d)$	0.5214
$VAR(d)$	0.0615

Table 4

Distribution of the nearest-neighbor distance from a given point in a unit square, assuming Poisson arrivals of N points

	Not considering edge effects	Including Donnelly (1978) correction terms
$f(d_1)$	$2\pi N d_1 e^{-\pi N d_1^2} : d_1 \geq 0$	–
$E(d_1)$	$\frac{1}{2\sqrt{N}}$	$\frac{1}{2\sqrt{N}} + \frac{1}{N} \left(0.0514 + \frac{0.041}{\sqrt{N}} \right)$
$VAR(d_1)$	$\frac{4-\pi}{4\pi N}$	$\frac{4-\pi}{4\pi N} + \frac{0.037}{N\sqrt{N}}$