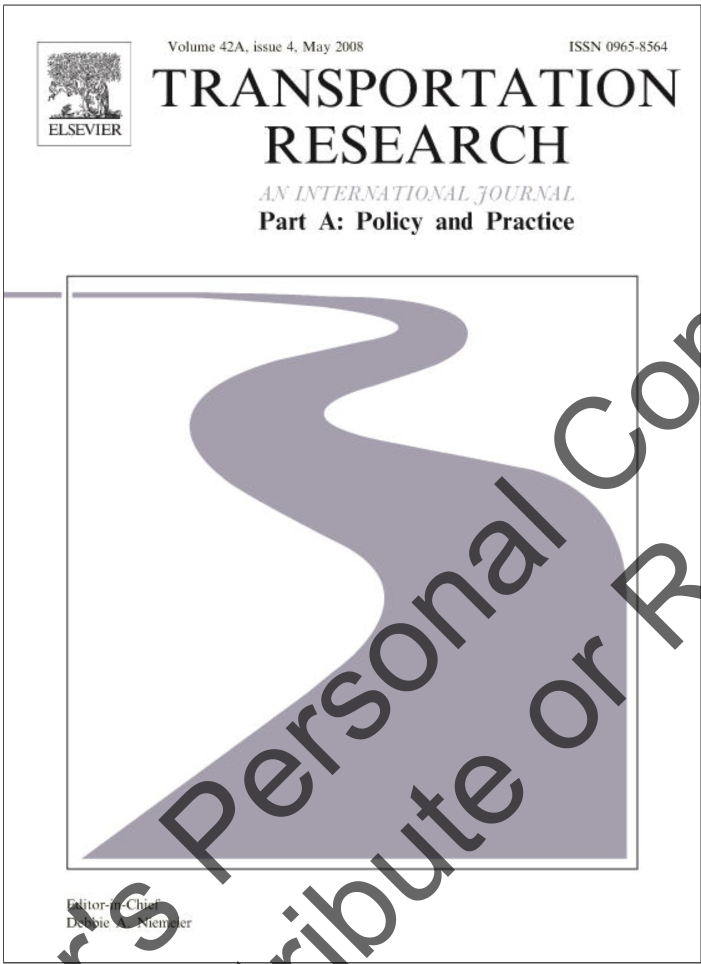
Volume 42A, issue 4, May 2008
ISSN 0965-8564
ELSEVIER
TRANSPORTATION
RESEARCH
AN INTERNATIONAL JOURNAL
Part A: Policy and Practice
Editor-in-Chief
Debbie A. Niemeier

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Transportation Research Part A 42 (2008) 718–737
TRANSPORTATION RESEARCH PART A
www.elsevier.com/locate/tra

A simulation study of demand responsive transit system design

Luca Quadrifoglio, Maged M. Dessouky *, Fernando Ordóñez

Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA 90089-0193, United States

Received 23 June 2005; accepted 1 January 2008

Abstract

In this paper we study the impact on productivity of specific operating practices currently used by demand responsive transit (DRT) providers. We investigate the effect of using a zoning vs. a no-zoning strategy and time-window settings on performance measures such as total trip miles, deadhead miles and fleet size. It is difficult to establish closed-form expressions to assess the impact on the performance measures of a specific zoning practice or time-window setting for a real transportation network. Thus, we conduct this study through a simulation model of the operations of DRT providers on a network based on data for DRT service in Los Angeles County. However, the methodology is quite general and applicable to any other service area. Our results suggest the existence of linear relationships between operating practices and performance measures. In particular we observe that for each minute increase in time-window size the service saves approximately 2 vehicles and 260 miles driven and that a no-zoning strategy is able to satisfy the same demand by employing 60 less vehicles and driving 10,000 less total miles with respect to the current zoning strategy.

Keywords: Dial-a-ride; Simulation; Time-windows; Zoning; Performance measures

1. Introduction

The passage of the Americans with Disabilities Act (ADA) has changed the landscape for demand responsive transit systems. First, the demand for this type of transit service has experienced tremendous growth. In Los Angeles County alone more than 5000 vans and 4200 cabs provide service, generating 8 million trips per year. Second, besides creating a larger demand, ADA also set strict guidelines for the providers on trip denials and on-time performance (Lewis et al., 1998). In essence, transit agencies today are expected to provide better services while experiencing increased usage for demand responsive transit systems.

The National Transit Summaries and Trends (NTST) report for 2002 indicates that the average cost per passenger trip for DRT systems is \$20.8 with fares ranging from \$1.5 to \$3.00. By way of contrast, the average cost per trip for fixed-route lines is \$2.4 with fares being roughly 25% of the cost. Therefore, DRT services are

* Corresponding author. Tel.: +1 213 740 4891; fax: +1 213 740 1120.
E-mail address: maged@usc.edu (M.M. Dessouky).

0965-8564/\$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tra.2008.01.018

still a highly subsidized service and it is imperative for agencies to analyze and investigate their current practices to identify possible cost reductions or productivity improvements.

To measure productivity and cost of the DRT system we consider different performance measures, such as fleet size, total miles and deadhead miles. The deadhead miles are defined as the *empty* trip miles driven by the vehicle between the drop-off point of a customer to the pick-up point of another customer. Note that with ridesharing a vehicle may not be empty driving from a drop-off point to a pick-up point and the miles driven in these cases would not count as deadhead miles. A reduction in deadhead miles can either cause a reduction in the total number of miles driven by a vehicle (hence reducing cost) or allow a vehicle to serve more customers on a given day (hence increasing productivity).

Some studies outline the potential positive impacts of Advanced Public Transportation Systems (APTS) on productivity and cost (Stone et al., 1994; Goeddel, 1996; Ben-Akiva et al., 1996; Chira-Chavala and Venter, 1997; Wallace, 1997; Schweiger and McGrane, 1999; Higgins et al., 2000; Stone et al., 2000). Palmer et al. (2004) show also how financial incentives and penalties can have a negative impact on productivity. That is, providers may schedule in an inefficient manner in order to ensure that they are on time to receive the incentive or avoid the penalty. But there are other factors that have an influence on the performance of DRT systems and the objective of this research is to study the impact on productivity and cost of specific operating practices currently used by DRT providers. They are the *time-window setting* and the *zoning*.

The length of the time-window that specifies the time range in which the provider must pick-up the customer is an important factor impacting productivity and cost. For example, a time-window of 20 min and a scheduled pick-up time of 3:00 pm would mean that the vehicle must pick-up the passenger by 3:20 pm at the latest to be considered on-time. Typically, providers have financial incentives or penalties for meeting on-time goals. Naturally, customers prefer small time-windows. However, in order to maintain small time-windows, transit agencies may have to decrease the ridesharing and increase their fleet size, contributing to increased cost and less productivity. Therefore, the setting of the time-window size needs to balance customer service with the impact on productivity and cost. Currently, Access Service Inc. (ASI), the agency responsible for coordinating paratransit DRT service in Los Angeles County, uses a 20 min time-window whereas many other agencies use a 30 min time-window.

A number of DRT agencies divide their service area into regions contracting the service in each of them to a different provider to simplify the management of the service. This practice, known as *zoning*, is also motivated by the drivers' preference to be assigned to a smaller region instead of the whole service area. This is a common practice for DRT agencies (paratransit, taxi services, etc.) all over the US especially when the service area is large. We distinguish between a *centralized* vs. *decentralized* control depending upon the number of regions in the service area. In centralized control, the service is aggregated into a single region; in decentralized control multiple regions are created. For example, ASI utilizes a decentralized control strategy dividing its service area into six regions (see Fig. 1). The pick-up location of the customer request determines the region and the corresponding provider responsible for the service. It is not uncommon that the pick-up and drop-off locations of a request are in different regions. In fact, according to the data provided by ASI, around 20% of the trips originating in the Northern region of Los Angeles County have a drop-off location outside that region. Hence, the return trip will be done by a different provider regardless of the dwell time of the customer at their drop-off location coming at the expense of a significant number of deadhead miles. Furthermore, in this situation, the customer is required to make two different reservations, one for each provider. In contrast, a hurdle toward implementing a more centralized strategy is that the Computer Aided Dispatching (CAD) systems of the different providers need to efficiently communicate among themselves in order to effectively manage such a design.

Although there is a significant body of work in the literature on scheduling and routing DRT systems (see e.g., Ioachim et al., 1995; Savelsbergh and Sol, 1995; Toth and Vigo, 1997; Borndörfer et al., 1999; Desaulniers et al., 2000; Diana and Dessouky, 2004; Lu and Dessouky, 2004, 2006), there has been no research performed comparing the performance of a centralized controlled DRT system with a decentralized one. Diana et al. (2005) developed analytical equations to determine the fleet size as a function of the time-window for a square service area. However, no similar equations exist for general service areas and for estimations of the total and deadhead miles. Thus, the effect of the time-window size on productivity and cost in general has also not been quantified. This paper addresses this gap by studying the impact of these issues on the operations of a representative large-scale DRT service.

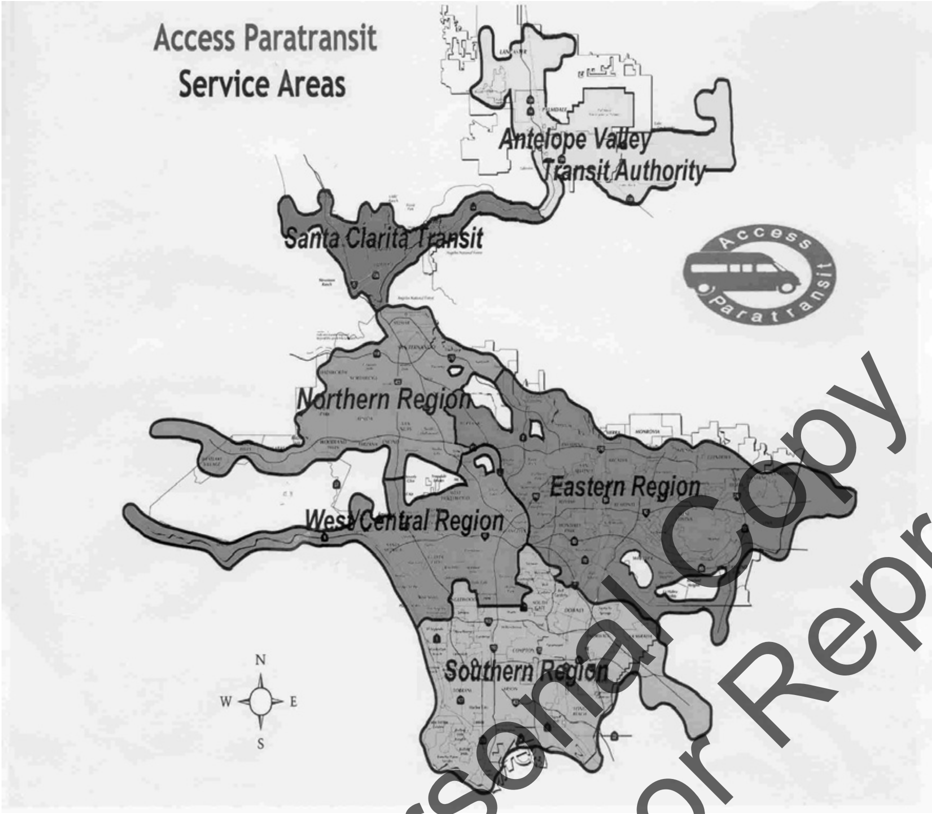

Fig. 1. Service regions in the Los Angeles County.

Due to the difficulty of developing closed-form expressions between the operating practices and the performance measures, a simulation model is used in this study. The simulation model is based on demand data provided by ASI for Los Angeles County. We analyze the effect of varying the time-window size (from 10 to 45 min) and we compare the current decentralized approach with a centralized strategy where any vehicle can pick-up any customer regardless of the service region. In addition we investigate the effect of centralizing only part of the service area, merging two regions together.

Although the results of the simulation model pertain to the Los Angeles County network, the simulation methodology described here is quite general. In fact, a similar study can be conducted for any DRT service with basic data (vehicle fleet, service parameters and description of demand), and therefore the methodology is easily adaptable and applicable to other service areas.

In addition, the results of this study provide insights on the dependency between performance measures and operating practices for DRT services in general. In fact, the geometry of the service area and the demand distribution of Los Angeles County utilized in this simulation model are quite standard as we describe later. Most urban areas (especially in the US) serviced by DRT systems would have similar configurations and we do not expect major differences on the nature of the relationships between the performance measures and operating practices for a number of other DRT providers.

Simulation tools are very powerful to evaluate systems' performance and they have been extensively utilized in the literature in a variety of fields including transportation. Wilson et al. (1970) pioneered the use of simulation to compare different heuristics to assess the influence of the service area, the demand density, and the service quality on the fleet size requirements. Regan et al. (1996) evaluate the performance of different load acceptance and assignment strategies for a dynamic distribution problem. Larson et al. (2002) examines the impact of dynamism on the quality of the solution for the Partially Dynamic Traveling Repairman Problem. Only a few applications are specifically concerned with paratransit systems. Fu (2002) develops a simulation model to assess the potential effects of the latest advances in information technologies on dial-a-ride paratransit systems. Deflorio et al. (2002) propose a simulation model to evaluate the performance of a DRT system scheduled using the insertion algorithm by Jaw et al. (1986) when dealing with random events like late customers and not on-time vehicles. Lipmann et al. (2002) and Hauptmeier et al. (2000) take traffic conditions into account to evaluate the performance of DRT systems. Some studies (Feuerstein and Stougie, 2001; Bailey and Clark, 1987) have investigated changes of performance when the dial-a-ride system is run with various numbers of vehicles. Haghani and Banihashemi (2002) address the relation between efficiency of vehicles and town size. Shinoda et al. (2004) compare by simulation the performance of dial-a-ride system vs. a fixed-route system in urban areas varying various parameters. Quadrifoglio and Dessouky (2007a) also perform a simulation study to test the efficiency of the insertion heuristic scheduling algorithm for Mobility Allowance Shuttle Transit (MAST) systems, a hybrid transit solution that merges the flexibility of DRT systems and the low cost operability of traditional fixed-route bus services. The same authors (Quadrifoglio and Dessouky, 2007b) use simulation to perform a sensitivity analysis of the performance of a MAST system varying the shape of its service area. Diana (2006) assesses by simulation the influence on the effectiveness of a DRT scheduling algorithm by Diana and Dessouky (2004) of typical dynamic parameters such as percentage of real time requests and interval between call-in time and requested pick-up time.

The remainder of the paper is organized as follows. In Section 2, we analyze the historical demand data of a representative large-scale agency (ASI). In Section 3 we describe the simulation model used to represent the Los Angeles County DRT system. This simulation model is described in sufficient detail so that it can serve as a source for simulations on other DRT system environments. Section 4 presents the results; finally the conclusions are outlined in Section 5.

2. Data analysis

In this section we conduct statistical analysis of DRT system demand data provided by Access Services Inc. (ASI), the agency designated to coordinate paratransit service in the Los Angeles County. Currently, local cities provide over 3.8 million annual paratransit trips throughout the county. ASI fills in the gaps and allows individuals to travel across the different regions within the service area of Los Angeles County. ASI has over 37,000 registered customers and divides the Los Angeles County in six regions; in each of them the paratransit service is contracted to a private operator; see Fig. 1. For the purpose of this study we consider only the Northern (N), Southern (S), Eastern (E) and West/Central (W) regions because Santa Clarita and Antelope Valley regions have very low demand compared to them (the daily average demand in each of these two regions is about 5% of the daily average demand in each of the four main regions). In addition, there are very few trips traveling from one of the four main regions to the Santa Clarita or Antelope Valley regions. Therefore we exclude them from the analysis.

In the following subsections we present in detail the real demand data provided by ASI consisting of requests originated in the Northern region covering 10 weekdays of service. We describe the demand data through the following demand features: number and type of passengers per request, call-in time, requested pick-up time, pick-up/drop-off locations and travel distance. The simulated demand will aim to match these features of the real demand.

2.1. Type of request

What defines the type of request are the number of passengers per trip and whether the customer(s) uses a wheelchair (W/C) or not (A = ambulatory). Each request can result in either a trip that is regularly scheduled and performed, a "No-Show" (the vehicle reaches the pick-up location but not the drop-off location because customers do not show up; however, both points needed to be scheduled) or a cancellation (customers cancel their reservation well in advance with no need of scheduling). There are a total of 12,842 requests in the 10 day period considered (from September 29th 2003 to October 10th 2003), not including the cancelled requests (about 20% of the total amount). Table 1 shows the daily average of requests with the standard deviation

Table 1
Daily requests, Northern region

Daily average # of requests	Std. dev.	W/C	No-Show
1284	5.6%	24.6%	9%

Table 2
Daily average performed trips, other regions

Region	Daily average # of trips (not including No-Shows)
West/Central	1328
Eastern	2009
Southern	1541

Table 3
Distribution of additional passengers

# of additional passengers	Request	
	A (%)	W/C (%)
0	86.22	65.15
1	13.32	31.56
2	0.45	3.07
3	0.01	0.22

(given as a percentage of the average), the percentage of W/C requests and the percentage of "No-Shows" for the Northern region.

ASI also provided the daily performed trips (not considering cancelled requests or "No-Shows") for the West/Central, Eastern and Southern regions, but without any further details. Table 2 shows those figures for the other regions.

The distribution of the number of additional passengers for A and W/C requests is shown in Table 3.

Furthermore, the probability that an additional passenger of a W/C request would also be on a wheelchair is 3.7%. This is important because wheelchair accessible vehicles have a limited number of seats for wheelchair passengers depending on their capacity.

2.2. Call-in time

Fig. 2 plots the distribution of the call-in time. The histogram shows that most calls occur in the morning with a peak between 6:00 am and 7:00 am; the frequency slowly decreases during the rest of the day with

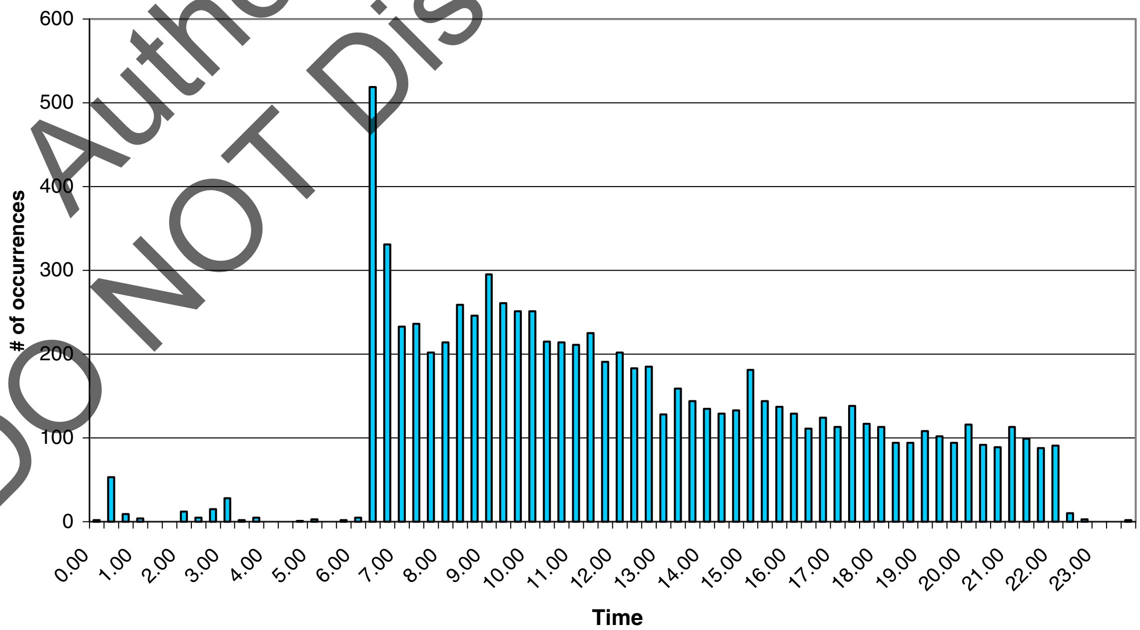

Fig. 2. Distribution of call-in time.

almost no calls during the night (10:00 pm-6:00 am). Several requests (about 35% of the total) did not have a valid recorded call-in time and have been excluded from this analysis. This is mostly due to automatic scheduling of recurrent requests equally repeated for several days, but booked only once (known as "standing" rides).

2.3. Requested pick-up time

The requested pick-up time distribution is shown in Fig. 3. The histogram clearly reveals that most of the rides are requested for daytime, between 6:00 am and 6:00 pm, with peaks between 8:00 am and 9:00 am and between 1:00 pm and 3:00 pm; a fewer amount of rides are requested for the evening and a very small percentage are needed for nighttime.

2.4. Interval between call-in time and requested pick-up time

The time interval between the call-in time and the requested pick-up time is shown in Fig. 4. The histogram shows that most requests are made well in advance with a time interval larger than 15 h. As noted earlier, a substantial amount of the requests (35%) did not have a valid recorded call-in time and have been excluded from this analysis. However, these requests are mostly the "standing" rides that are booked once and are valid for several days, therefore made well in advance too. We also note that the current policy of ASI guarantees scheduling of the service for all the requests made the day before or earlier; while the reservations made the same day are not guaranteed, but they may be accommodated into the current schedule when possible.

2.5. Pick-up and drop-off locations

The next figures show the distribution of the pick-up and drop-off locations. In both Figs. 5 and 6 each square represents a one-square mile area.

Since the demand analyzed corresponds to the demand faced by the Northern region, the pick-up locations (Fig. 5) are all inside the Northern region (see Fig. 1). The drop-off locations (Fig. 6) are distributed along the whole Los Angeles County area as described also in Table 4.

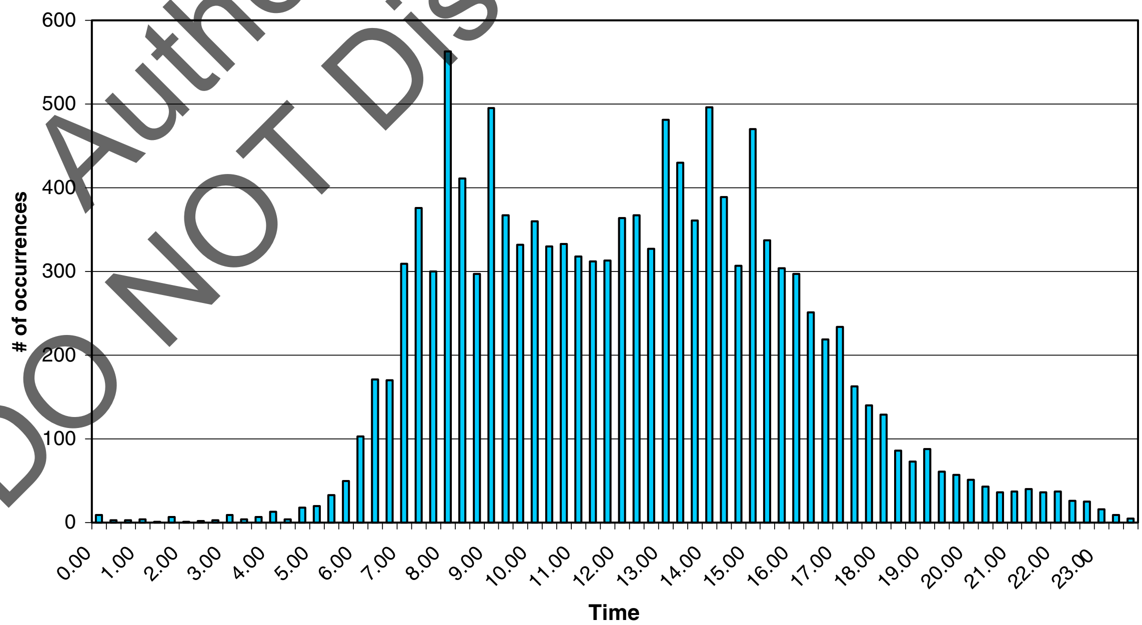

Fig. 3. Distribution of requested pick-up time.

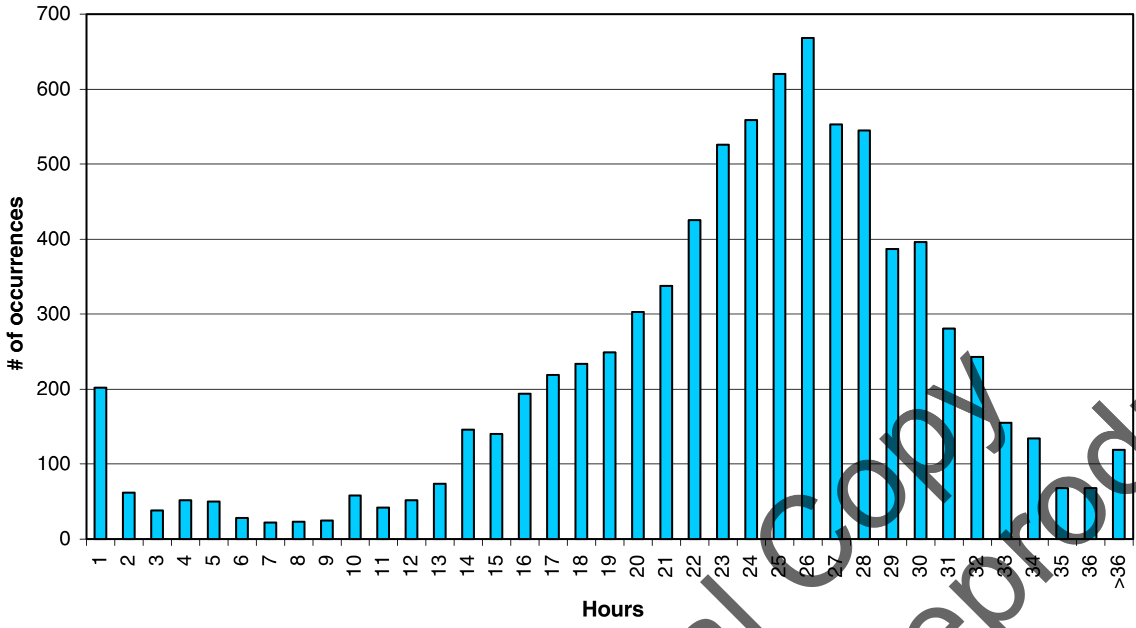

Fig. 4. Distribution of time interval call-in/pick-up.

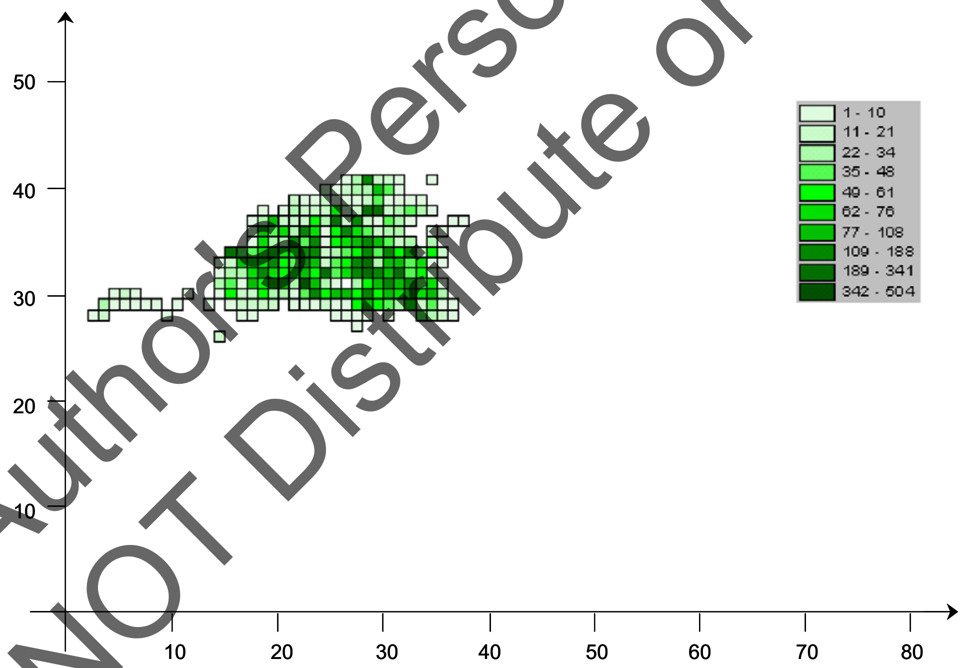

Fig. 5. Distribution of pick-up locations (miles).

2.6. Travel distance

Fig. 7 shows the distribution of the travel distance between pick-up and drop-off locations. Most of the trips have a distance not larger than 15 miles as most of the trips are within the pick-up region. Statistics about the distribution are provided in Table 5.

The distances for each request are provided by ASI and they are computed by geocoding pick-up and drop-off locations. We noticed that in most cases the actual travel distance closely matches the rectilinear distances calculated between the pick-up and drop-off locations of each request.

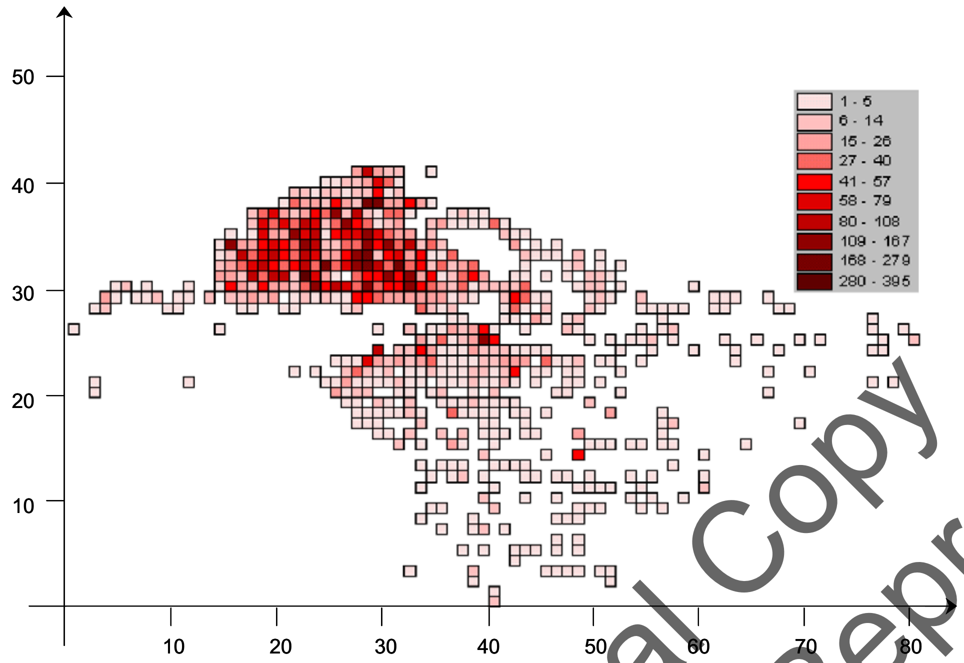

Fig. 6. Distribution of drop-off locations (miles).

Table 4
Distribution of pick-up and drop-off location by zone

Region	Pick-ups	%	Drop-offs	%
Northern	12,842	100	10,204	79.5
Outside	0	0	2,638	20.5
West/Central	0	0	1,320	10.3
Eastern	0	0	1,028	8
Southern	0	0	290	2.2

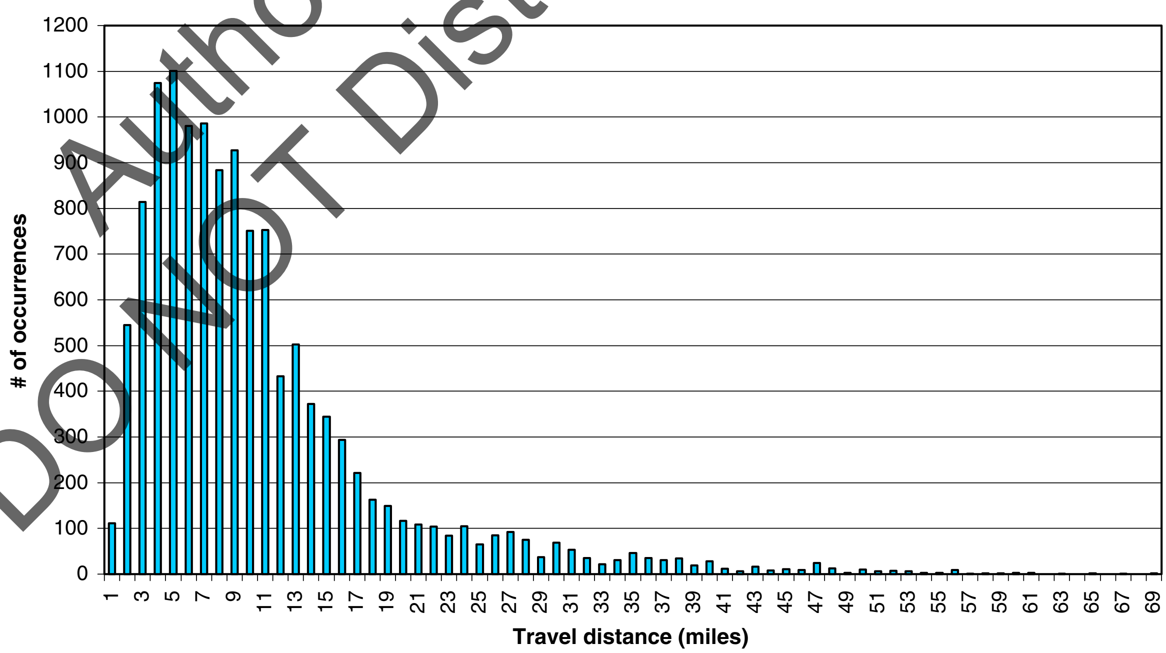

Fig. 7. Distribution of travel distance

Table 5
Travel distance statistics

Average	10.2
Median	7.9
Standard deviation	8.4
Minimum	0.5
Maximum	68.8

3. Simulation model

In this section we explain how we generate random samples to emulate the demand, based on the distributions found in Section 2 in order to develop a simulation model that is representative of the area serviced by ASI. We also describe the fleet size for each region, the service parameters used, and the scheduling algorithm.

Our demand simulation assumptions can be separated into two categories:

1. Assumptions required to address the fact that we lack detailed data for Los Angeles County area except for the Northern region. These assumptions allow us to create a representative network which although is not exactly the Los Angeles County DRT system, it closely resembles the key features of it. We will discuss them in Section 3.1.
2. Assumptions needed to generate demand distributions that resemble the observed real demand features. We will cover these issues in Section 3.2.

3.1. Network assumptions

The assumptions made to circumvent the lack of demand data on the whole LA County are particular to this study and are geared to obtain the missing demand information. The ideal situation would be to obtain real demand data for the entire region and avoid these approximations.

The key assumption is to consider the features of the demand distribution in each region be the same as the detailed demand information we have for the Northern region with a few caveats. For example, total daily demand in each region is the corresponding average daily demand shown in Table 2. In addition, the geographical shape of each region is carefully replicated as in Fig. 1 and the distributions of the pick-up and drop-off locations are dependent on each region.

Note that the four regions are all adjacent to each other except the Northern and Southern regions. We will use this adjacency property to make assumptions about the proportion of out-of-zone drop-offs.

3.1.1. Daily requests per region

We assume that the total number of requests including "No-Shows", but not cancelled requests, in a given simulated day is the daily average (rounded to the nearest hundred) according to the data provided by ASI. Thus, for the Northern region we have 1300 customers/day rounded from the actual value of 1284 (Table 1). For the other regions the numbers of daily requests are obtained by first adding to the numbers in Table 2 an assumed 10% of "No-Show" requests (rounded from the actual 9% for the Northern region, see Table 1) and then rounding these figures to the nearest hundred. [i.e.: for the Eastern region we have 2009 actual performed trips; we add a 10% of "No-Show" obtaining $2009/0.9 \approx 2232$, rounded to 2200].

Table 6 summarizes those figures.

Table 6
Assumed number of daily requests

Region	Daily requests (including "No-Shows")
Northern	1300
West/Central	1500
Eastern	2200
Southern	1700

3.1.2. Pick-up/drop-off locations per region

We derive the pick-up and drop-off locations distributions for regions W, E, and S by extrapolating the results from the Northern region in the following way.

The demand data in Figs. 5 and 6 show that the pick-up locations distribution and the in-zone drop-off locations distribution for the Northern region are very similar. We thus infer that in each region the pick-up and the drop-off locations distributions resemble each other closely. Under this assumption we create the pick-up location geographical distributions for the other three regions using as a template the out-of-zone drop-off locations distribution from the data of the Northern region.

The drop-off locations distributions used when simulating the W, E and S regions are the same as the ones used for the Northern region.

3.1.3. In-zone/out-of-zone drop-off percentage

Finally we need to determine the percentage of requests having in-zone and out-of-zone drop-offs. For this we use the adjacency property mentioned earlier: from the probabilities shown in Table 4 for the N region we obtain the out-of-zone drop-off percentages in an adjacent region (W and E: 10.3% and 8% rounded to 10%) and a non-adjacent region (S: 2.2% rounded to 5%). That is, we assume that the probability that a request would have the drop-off location in an adjacent region is 10%, while the probability of having a drop-off location in a non-adjacent region is 5%; the in-zone drop-off percentages for each region are the balance. Table 7 summarizes the values used for the probability of the drop-off region for each pick-up location region Table 8.

3.2. Demand assumptions

We now describe the simulation assumptions we make to generate demands that behave similar to the historic data on requests for service.

3.2.1. Type of request

We decide by generating random numbers whether each request requires a W/C accessible vehicle (with a probability of 25%). Then, we determine how many extra passengers there are based on the probabilities shown in the following tables.

Furthermore, if the request requires a W/C accessible vehicle and there are extra passengers, we decide whether each extra passenger is a W/C as well with a probability of 5%. We also compute (with a probability of 10%) whether a request results in a "No-Show".

Table 7
Drop-off region probabilities

Pick-up	Drop-off			
	Northern (%)	West/Central (%)	Eastern (%)	Southern (%)
Northern	75	10	10	5
West/Central	10	70	10	10
Eastern	10	10	70	10
Southern	5	10	10	75

Table 8
Probability of having additional passengers

Request	# of additional passengers	
	A (%)	W/C (%)
0	85	65
1	13	30
2	2	4
3	0	1

Note that the above probabilities are rounded values based on the sample probabilities in Section 2.

3.2.2. Call-in time

Since most of the requests are made well in advance (Fig. 4), we assume a static environment and we suppose that all the requests for any given day of service are made at least one day in advance.

3.2.3. Requested pick-up time

In order to simulate the requested pick-up time for each request, we use a piecewise linear approximation of the actual cumulative distribution obtained by the data shown in Fig. 3. Fig. 8 shows this actual cumulative distribution (dotted line) of the requested pick-up time and the piecewise linear approximation (solid line) used in our simulation. Most of the actual pick-up requests occur between 6:00 am and 6:00 pm. The piecewise linear approximation assumes that 94% of the pick-up requests occur between 6:00 am and 6:00 pm; only 6% of them occur before (3%) or after (3%) this interval. As a result, the two cumulative distributions are very similar to each other. Once a requested pick-up time is sampled, we rounded its value to the nearest multiple of 5 min (for example: 6:00, 6:05, 6:10, etc.).

3.2.4. Pick-up/drop-off locations

The methodology to generate pick-up/drop-off locations must take into consideration two features of the demand: the actual locations visited by the vehicles to pick-up and drop-off customers and the travel distance of the trips. In fact, they all directly impact the vehicle scheduling and consequently the performance measures we are interested in (e.g., miles traveled, deadhead miles).

Even though we have detailed data from ASI for the Northern region about pick-up and drop-off locations and travel distance distributions, their generation in the simulation model is not so simple because these three variables depend on each other. For instance, if we sample pick-up and drop-off locations from their distributions, the travel distance would be calculated as a dependent variable and could not be sampled from the actual distribution (Fig. 7). If pick-up and drop-off location distributions were to be statistically independent, then the calculated travel distance distribution could match the actual one exactly. But this may not be the case due to the following dependencies among pick-up and drop-of locations of each request:

I. Customers typically do not require a ride having the drop-off point very close to the pick-up point.
II. When customers require a drop-off (pick-up) at high demand density locations (such as hospitals, malls, schools, parks, etc.) they would most likely choose the ones closest to their pick-up (drop-off) location (home, office, etc.).
III. An existing fair amount of recurrent customers (such as the "standing" rides) or standard itineraries creates strong links between some particular pick-up/drop-off pairs.

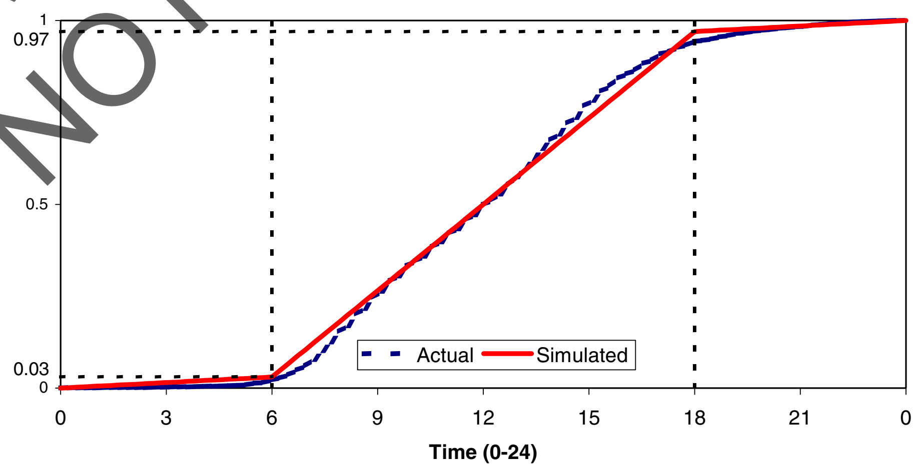

Fig. 8. Actual and simulated cumulative distribution of requested pick-up time.

The actual travel distance distribution is the result of the correlation between pick-up and drop-off locations explained by all the factors above (and possibly many others). Ideally we would need to generate a different drop-off (or pick-up) location distribution for each pick-up (or drop-off) point in order to capture their actual dependency and replicate the actual demand in our simulation model. Clearly this procedure would be very hard to implement.

Therefore, in order to generate the pick-up and drop-off locations for each request we considered two methods and we evaluate them by comparing the resulting simulated distributions with the actual ones.

- Option 1: Independently sample pick-up and drop-off locations from the distributions in Fig. 5 and 6.
- Option 2: Separate demand data into four groups depending on the region of the drop-off location (N, E, W or S), thus obtaining four different pick-up location distributions. Sample first the drop-off region visited based on probabilities from Table 4. Then sample pick-up and drop-off locations from the demand group selected.

We also considered investigating a third option. This procedure samples the pick-up from Fig. 5 and the drop-off from Fig. 6, but using them only to determine trip pick-up and direction. The travel distance is sampled from Fig. 7. This method will perfectly replicate the travel distance distribution while sacrificing the accuracy of the drop-off locations, which would be calculated as a dependent variable and could end up outside of the regions. Since the main focus of this research is to analyze the impact of the deadhead miles mainly resulting from travel distances between an out-of-zone drop-off and the next in-zone pick-up, it is crucial that the drop-off locations (especially the out-of-zone ones) are correctly sampled. Therefore we chose not to pursue this third option.

Option 1 is the simplest sampling method, but it ignores potential shortcomings because of the above I–III. Option 2 attempts to better represent the demand trying to capture their mutual dependency by separating it into four groups. The issue that we want to investigate is how much Option 1 differs from the actual data and how much Option 2 better represents the real demand compared to Option 1.

Fig. 9 shows a comparison between the actual travel distance distributions and the distributions generated by Option 1 and Option 2. The statistics of the distributions are shown in Table 9.

Fig. 9 and Table 9 show that the difference between Option 1 and the actual distribution is not very large overall. However, the discrepancies are explained by the above points I–III. In fact, the actual distribution does not have values around zero (point I). The actual distribution is consistently more skewed towards

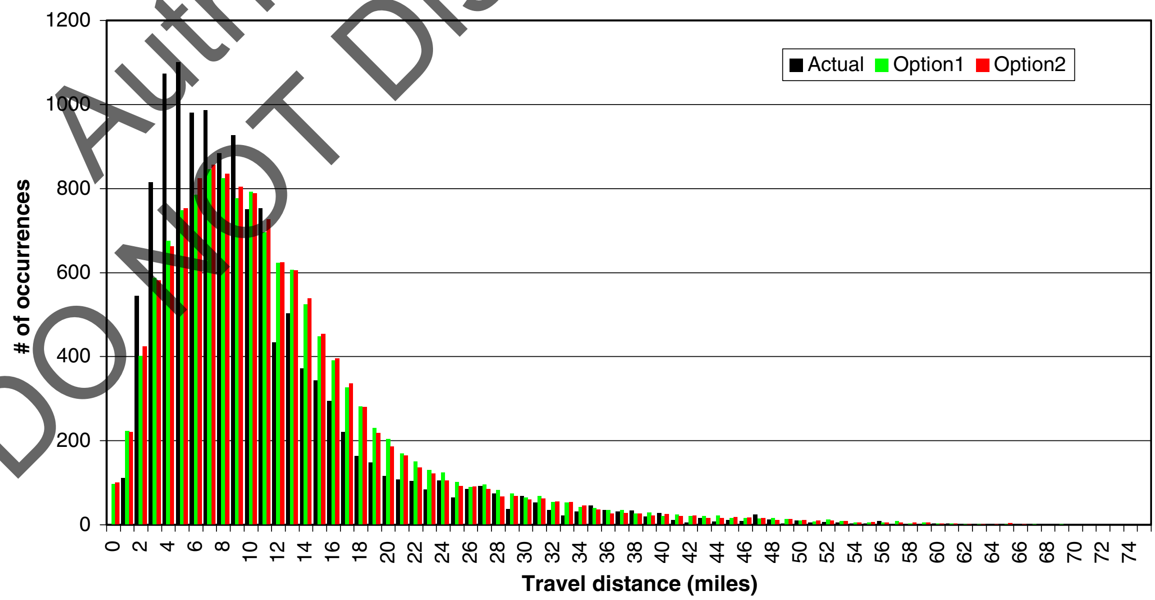

Fig. 9. Comparison between actual and simulated travel distance distributions.

Table 9
Travel distance distribution statistics (aggregate)

	Actual	Option 1	Option 2
Average	10.2	12.24	12.03
Median	7.9	10	10
Standard deviation	8.4	9.21	9.05
Minimum	0.5	0	0
Maximum	68.8	73	74

shorter values (point II). Finally, the presence of regular customers (point III), although is not observed in the cumulative distribution, can be seen from the demand data provided by ASI, where some "typical" trips are consistently repeated ("standing" rides).

Furthermore, the chart and the table indicate that Option 1 and Option 2 produce very similar results. In fact, even though Option 2 is slightly closer to the actual distribution, the improvement over Option 1 is so small that it does not justify the demand grouping. This result points out that there is no significant dependency between the pick-up locations distributions and the drop-off region. The demand would need to be divided in a much larger amount of groups (not only four) in order to observe a significant improvement and a closer match between the simulated and the actual distributions. But since Option 1 does not differ very much from the actual distribution to begin with, the marginal improvement in quality of the generated demand does not justify the additional computational burden.

Thus, since Option 1 and the actual distribution do not differ very much and the improvement obtained by Option 2 is not significant, we select Option 1 as a sampling methodology for pick-up and drop-off locations assuming therefore independency.

In order to incorporate the assumptions discussed in Section 3.1 as well, the sampling procedure is therefore the following: for each region and for each request we first sample the pick-up location from its distribution in that region; then by generating a random number based on Table 7 we decide in which region the drop-off point will be located and we sample it from the corresponding drop-off locations distribution.

3.3. Resources

There are different vehicle types owned by ASI and used by the local operators to provide the service depending on the seat capacity for ambulatory (A) and wheelchair (W/C) passengers. ASI provided its current fleet capacity for each region and type. This is summarized in Table 10.

ASI also provided the locations of the four depots (one for each region). Each private operator owns spare vehicles (not included in Table 10) in order to assure the service in case of need during high demand periods.

3.4. Service parameters

The parameters used in the simulation are the following:

- Vehicles' average speed: 25 miles/h.
- Service time at each boarding/disembarkment if there are no W/C customers: 30 s.

Table 10
Fleet: # of vehicles of each type for each region

Region	Vehicle type				
	3 A + 2 W/C	8 A + 2 W/C	8 A + 3 W/C	11 A + 6 W/C	7 A + 0 W/C
Northern	100	3	10	2	15
West/Central	88	12	1	1	0
Eastern	123	7	0	0	0
Southern	61	11	23	2	0

- Service time at each boarding/disembarkment if there is at least one W/C customer: 10 min.
- Maximum ride time factor: 2.5. That is, the actual ride time from pick-up to drop-off points for each request should not exceed 2.5 times the direct ride time.
- Time-windows: 20 min as a base case (as provided by ASI). As mentioned, we will perform sensitivity analysis over this parameter (Section 4.1).

3.5. Scheduling algorithm

After sampling the daily demand we utilized a sequential insertion algorithm to schedule the trips by assigning customers to vehicles according to the available fleet and the time-window constraints. Insertion algorithms are common procedures used by DRT providers to schedule their service, as noted by Lu and Dessouky (2006) and Campbell and Savelsbergh (2004) because of their computational efficiency. ASI and its regional providers conform to this common practice.

We imposed a maximum total shift length of 12 h. Most of the shifts run approximately from 6:00 am to 6:00 pm. In order to satisfy the nighttime demand, a few vehicles are scheduled for double shifts (with two drivers): approximately from 12:00 am to 12:00 pm and from 12:00 pm to 12:00 am. These criteria are consistent with the data provided by ASI for the Northern region and assumed to be the same for all four regions.

Vehicles begin and end their shifts at the depot of the corresponding region. For each depot the algorithm fills up one vehicle at a time, starting with the ones with bigger capacity currently available. The heuristic inserts in the current schedule the request, among the available and feasible ones, that minimizes the additional distance to be traveled. This procedure is carried out iteratively until the vehicle cannot accept any more requests because of time-window, capacity or time constraints. Then the algorithm checks if smaller capacity available vehicles could be used for the schedule just built and, if so, the smaller capacity vehicle is assigned. The vehicle does not have an assigned shift a priori, but this is determined during the insertion procedure: when and if an inserted request forces the schedule to begin before 6:00 am or to end after 6:00 pm, then this vehicle is used for a double shift (12:00 am/12:00 pm and 12:00 pm/12:00 am); otherwise the vehicle is used for a daytime shift (6:00 am/6:00 pm). Additional vehicles are filled up until all the requests have been scheduled.

4. Results

In this section we describe the results obtained by our simulation analysis. The performance measures used to evaluate the findings are the total number of vehicles used to satisfy the demand, the total miles driven, the deadhead miles driven, the passenger-miles and idle time. The deadhead miles represent miles driven by the vehicles with no passengers onboard; the idle time is the total time spent by the vehicles while waiting idle for the next scheduled pick-up. We illustrate the time-window size effect in Section 4.1 and the zoning effect in Section 4.2.

4.1. Time-window size effect

This section quantifies how much the performance measures are affected by variations of the time-window size. We performed again 30 replications of one day of service. For each day we schedule the service varying the time-window size from 10 to 45 min, with steps of 5 min. In the table we also include the figures for the asymptotic cases time-window settings of 0 and ∞. For all these cases we always assumed a decentralized strategy (zoning). The results are summarized in Table 11 in terms of averages and standard deviations given as a percentage of the average.

The effect of widening the time-window size is evident in all the performance measures: lower number of used vehicles, less total miles, less deadhead miles driven, less idle time spent, and increased passenger-miles because more ridesharing is possible. Larger time-windows imply looser time constraints and a larger feasible solution set. Thus more feasible options are available for the algorithm when building the schedules and the system can be solved more efficiently. Of course all these improvements come at a cost for the customers that have to deal with undesired larger time-windows.

Table 11
Time-window effect

Time-window minutes	# of vehicles	Std. dev. (%)	Total miles	Std. dev. (%)	Deadhead miles	Std. dev. (%)	Passenger-miles	Std. dev. (%)	Idle time (min)	Std. dev. (%)
0	614	0.9	125,400	0.6	44.096	1.1	172,060	0.9	97,319	1.8
5	561	1.1	120,410	0.8	39,769	1.6	180,621	0.9	72,469	2.2
10	531	1.4	117,107	1.0	37,234	2.0	185,611	0.9	59,328	1.9
15	511	1.3	114,739	0.9	35,470	1.5	189,769	0.8	50,905	2.8
20	497	1.1	113,021	0.9	34,131	1.6	193,011	0.8	45,397	3.1
25	485	1.1	111,650	0.8	33,013	1.6	196,220	0.7	41,479	2.9
30	474	1.2	110,308	0.9	32,076	1.7	198,386	0.6	38,346	3.0
35	468	1.2	109,445	0.8	31,414	1.5	200,415	0.8	35,842	3.3
40	459	1.3	108,514	0.9	30,755	1.9	202,674	0.8	33,484	3.4
45	452	1.2	107,532	0.9	29,952	1.7	205,062	0.9	31,508	3.3
…	…	…	…	…	…	…	…	…	…	…
∞	252	0.7	87,103	0.6	18,133	0.4	286,149	1.1	7,723	0.8

The values in Table 11 show monotonic behaviors for all the performance measure. They begin with a specific value at the origin (time-window set to zero) and they reach asymptotic values for infinite time-windows. Their curves can therefore be approximated by exponential types of equations (such as $b \pm ae^{-x}$). However, in the operating interval considered (time-window from 10 to 45 min) the values show almost linear relationships between the independent variable (time-window size) and the dependent variables (performance measures). In Figs. 10–14 we plot the mean values of performance measures vs. time-window size and the linear regression model for each of them. In each case we also present the R^2 and the p-value corresponding to the slope of the relationship.

In all the charts we note that the mean values seem to be part of a slightly convex (or concave for passenger-miles) curves, that are reasonably well approximated by linear regressions in this range, as indicated by the high R^2 values. The small p-values for the slope indicate that the probability that there is a relationship between each performance measure and time-window size is greater than 99% in all five cases. Assuming that the linear regressions can be used as good estimates, we are able to predict the savings for each performance measure per every minute added to the time-window size. From the regression formulas we have approximately:

- 2 vehicles less to be employed.
- 260 total miles less to be driven.
- 200 deadhead miles less to be driven.
- 530 increase in the passenger-miles value.
- 750 idle minutes less to be waited.

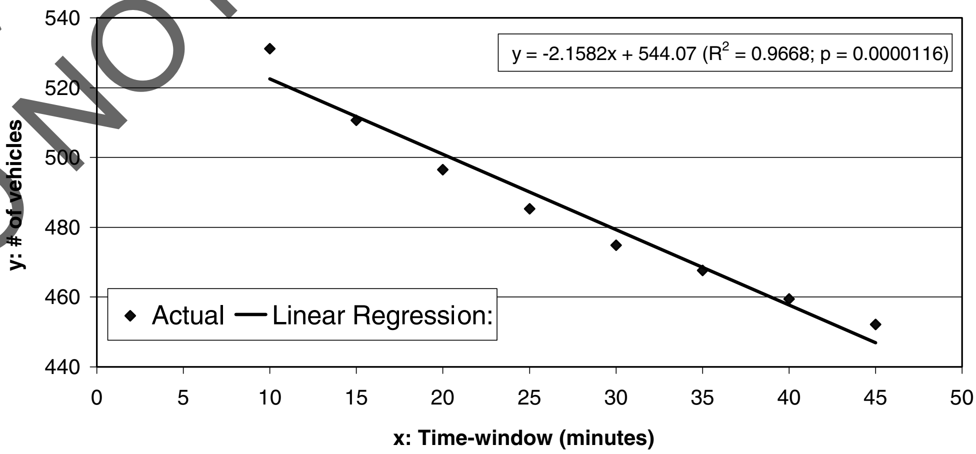

Fig. 10. Linear regression: number of vehicles.

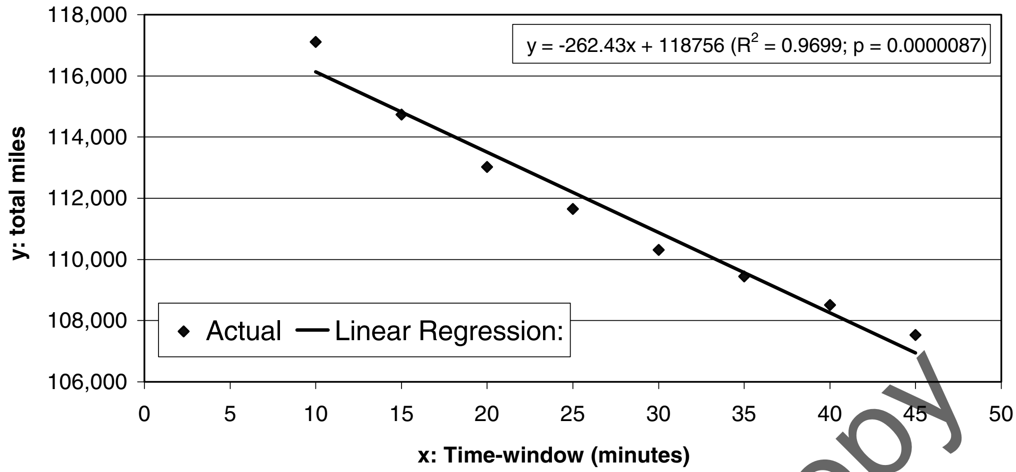

Fig. 11. Linear regression: total miles.

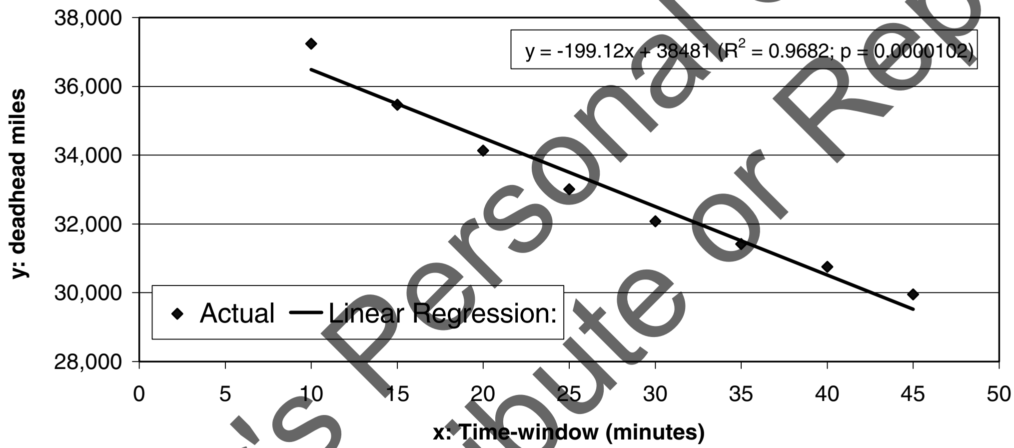

Fig. 12. Linear regression: deadhead miles.

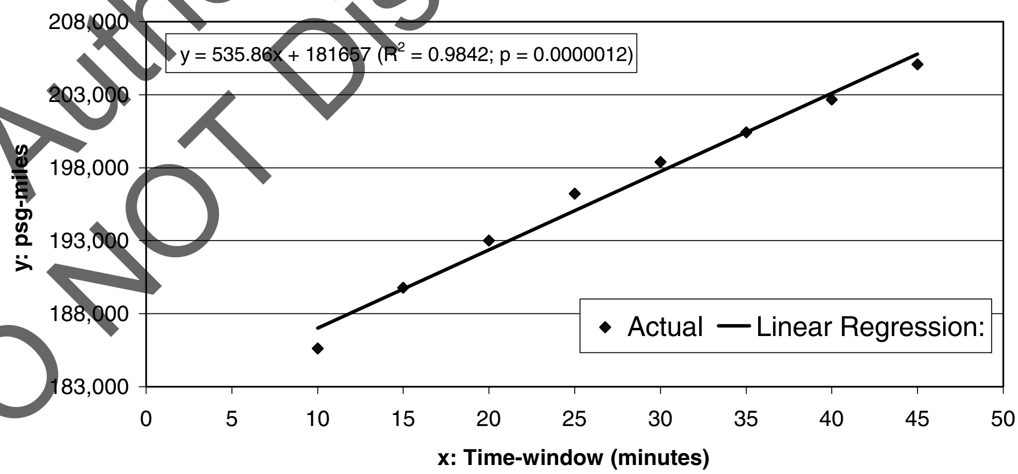

Fig. 13. Linear regression: passenger-miles.

The above figures can be used by managers when deciding the time-window size of the service. The impact on productivity and cost needs to balance a proper customer service. If ASI were to increase the time-window

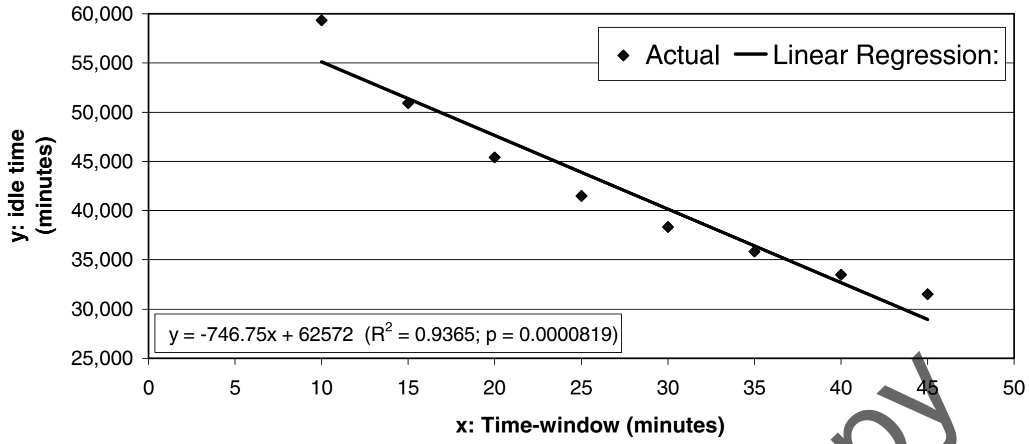

Fig. 14. Linear regression: idle time (minutes).

size from the current 20 min to 30 min, the improvement will be a saving of 20–25 vehicles and about 2600 miles less to be driven (2000 of which would be deadhead miles).

4.2. Zoning effect

The effect of a centralized strategy (zoning) vs. a decentralized strategy (no-zoning) is investigated in this section. The variations in productivity and costs are measured in terms of the performance measures defined above. We performed this analysis using a time-window size of 20 which is the current size that is being used by ASI.

We performed 30 replications of one day of service. For each day we schedule the service with four different policies:

1. Zoning (N/W/E/S): each zone takes care of its own pick-up requests (the current practice of ASI). This is the decentralized strategy.
2. Partial zoning (NW/ES): zones are merged together in groups of two regions and the requests are scheduled considering these two new formed larger regions.
3. Partial zoning (NE/WS): analogous to point 2, but different grouping.
4. No-zoning (NWES): the Los Angeles area is considered as a unique region. This is the centralized strategy.

We need to specify that the scheduling algorithm in case of no-zoning or partial zoning is slightly modified because we have more than one depot to choose from. Therefore, before scheduling a new vehicle the algorithm decides which depot to consider by selecting the one with the minimum distance (depot → pick-up → drop-off → depot) to its closest available request. Then the algorithm proceeds as explained in Section 3.5.

Table 12 summarizes the results in terms of the averages and standard deviations which are given as a percentage of the average of the data from the 30 replications.

Table 12
Zoning effect

Zoning	# of vehicles	Std. dev. (%)	Total miles	Std. dev. (%)	Deadhead miles	Std. dev. (%)	Passenger-miles	Std. dev. (%)	Idle time (min)	Std. dev. (%)
N/W/E/S	497	1.1	113,021	0.9	34,131	1.6	193,011	0.8	45,397	3.1
NW/ES	469	1.2	109,388	1.0	30,431	2.0	193,931	0.6	42,281	3.9
NE/WS	461	1.2	107,358	0.9	28,755	1.8	193,896	0.8	42,733	2.4
NWES	437	1.1	102,675	0.8	24,701	2.3	194,165	0.8	39,727	3.4

Table 13
Gaps obtained by doubling the number of zones

Policy	# of vehicles	Total miles	Deadhead miles	Passenger-miles	idle time (min)
a-Zoning (4 zones)	497	113,021	34,131	193,011	45,397
b-Partial zoning (2 zones) (NW/ES-NE/WS average)	465	108,373	29,593	193,914	42,507
c-No-zoning (1 zone)	437	102,675	24,701	194,165	39,727
Gap b → a	32	4,648	4,538	−903	2,890
Gap c → b	28	5,698	4,892	−251	2,780

Comparing the results obtained from the zoning case (N/W/E/S) and the no-zoning case (NWES) the savings are evident in terms of deadhead miles and number of vehicles used to serve the same demand. The NW/ES and NE/WS cases show intermediate outcomes and they are similar to each other. The cost of zoning vs. no-zoning is of about 60 extra vehicles to be employed that are needed because of the necessary extra deadhead miles (10,000) to be driven from the out-of-zone drop-off points to the next in-zone pick-up point. By allowing a no-zoning policy, many of these deadhead miles are no longer necessary because the next pick-up point does not necessarily have to be located in-zone, but it can be located anywhere in the Los Angeles area and the scheduling algorithm would be able to choose a closer one to the last drop-off point.

The differences in the total miles driven among the four cases are mostly due to the differences in the deadhead miles. In fact, subtracting the deadhead miles from the total miles driven in all cases we obtain roughly the same numbers. However we do observe a slight improvement in these numbers when comparing no-zoning vs. zoning. Therefore the zoning effect, besides the obvious saving in deadhead miles, affects also the efficiency of the scheduling algorithm itself, even if very slightly. Because without zoning constraints more feasible insertions (requests belonging to any region) are available to be chosen by the algorithm when scheduling a vehicle; thus the feasible solution set is bigger and better solutions could be found. This larger "freedom" in choosing the insertions allows the algorithm to fill up the vehicles' schedules more densely, explaining the reduction in idle time, and with more ridesharing as shown by the slight increase in passenger-miles.

The drastic reduction of the deadhead miles and the slight improvement of the algorithm efficiency explain the improvements in the figures. The first effect is by far the most significant one and its importance is directly proportional to the amount of out-of-zone drop-off points of the demand data; in this case about 20% of the total demand (see Table 4). Clearly, with no out-of-zone drop-off points there will be no savings in deadhead miles when switching from a zoning to a no-zoning policy. Conversely, the higher the out-of-zone drop-off points percentage, the more significant would be the improvement.

As for the time-window size setting analysis, we can infer a linear dependency between the performance measures and the zoning choice. To do so, we average together the results of the partial zoning cases (NW/ES and NE/WS) and we compute the gaps in each performance measure value while doubling the number of zones, namely going from a no-zoning to a partial zoning policy and from a partial zoning to a zoning policy as shown in Table 13.

The obtained gaps (b → a and c → b) for each performance measure are very similar showing again almost linear relationships between the zoning choice and the performance measures. Only the passenger-miles gaps are not comparable to each other, but they are also very small compared to the mean values showing a very low dependency on the zoning choice.

The clear advantages of the no-zoning policy shown by the figures above need to be carefully balanced with the added complexity of managing a larger service area. The local providers would need to closely work together. In particular the Computer Aided Dispatching (CAD) systems of different providers would need to efficiently communicate among themselves in order to effectively manage the service.

5. Conclusions

In this paper we quantified how much productivity and cost of Demand Responsive Transit (DRT) services are affected by two managerial practices: the time-window size setting and a centralized vs. decentralized strategy. Access Services Inc. (ASI), which is the designated consolidated transportation service agency to coordinate paratransit service within the Los Angeles County, provided us with demand data to generate statistical distributions used for a simulation model for our analyses. Although the results pertain to the network considered, the simulation methodology described here is quite general and easily applicable to any other large service area.

The results of this study provide general insights on the dependency between performance measures and operating practices for large DRT services. We identified quasi-linear relationships between the performance measures and the independent variable, either the time-window size or the zoning policy.

For the time-window size effect, we built linear regression models and observed that for each minute increased in the time-window size the service saves approximately 2 vehicles and 260 miles driven, while satisfying the same demand. Increasing the time-window size would also lower the customer satisfaction, thus managers have to carefully balance these two effects while setting the size.

About the zoning policy, we observed that a centralized strategy is able to satisfy the same demand by employing 60 less vehicles and driving 10,000 less total miles with respect to a decentralized strategy. Most of the improvement is due to the drastic reduction of deadhead miles driven in the no-zoning (centralized) case. This increased efficiency has to be carefully balanced with the added complexity arising while managing centralized systems.

Future research would include using approximations in order to build analytical models to quantify the zoning and the time-window size effects and compare their findings with the ones from this simulation analysis.

Acknowledgement

This research was entirely supported by PATH and Caltrans. The research team greatly appreciates the continuous support of Access Services, Inc. (ASI) for providing both technical expertise and data in this research.

References

Bailey Jr., W.A., Clark Jr., T.D., 1987. A simulation analysis of demand and fleet size effects on taxicab service rates. Proceedings of the 19th Conference on Winter Simulation. ACM Press, pp. 838–844.

Ben-Akiva, M., Benjamin, J., Lauprete, J., Polydoropoulou, A., 1996. Impact of advanced public transportation systems on travel by dial-a-ride. Transportation Research Record 1557, 72–79.

Borndörfer, R., Grötschel, M., Klosteimer, F., Küttner, C., 1999. Telebus Berlin: vehicle scheduling in a dial-a-ride system. Computer-Aided Transit Scheduling. In: Lecture Notes in Economics and Mathematical Systems, vol. 471. Springer-Verlag, Berlin, pp. 391–422.

Campbell, A.M., Savelsbergh, M., 2004. Efficient insertion heuristics for vehicle routing and scheduling problems. Transportation Science 38, 369–378.

Chira-Chavala, R., Venter, C., 1997. Cost and productivity impacts of a smart paratransit system. Transportation Research Record 1571, 81–87.

Deflorio, F.P., Dalla Chiara, B., Murro, A., 2002. Simulation and performance of DRTS in a realistic environment. In: Proceedings of the 9th Meeting EWGT on "Intermodality, Sustainability and Intelligent Transportation Systems" and 13th Mini EURO Conference on "Handling Uncertainty in the Analysis of Traffic and Transportation Systems", Bari, Italy.

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

Diana, M., 2006. The importance of information flows temporal attributes for the efficient scheduling of dynamic demand responsive transport services. Journal of Advanced Transportation 40, 23–46.

Diana, M., Dessouky, M., 2004. A new regret insertion heuristic for solving large-scale dial-a-ride problems with time windows. Transportation Research B 38B, 539–557.

Diana, M., Dessouky, M.M., Xia, N., 2005. A model for the fleet sizing of demand responsive transportation services with time windows. Transportation Research, Part B: Methodological 40, 651–666.

Feuerstein, E., Stougie, L., 2001. On-line single-server dial-a-ride problems. Theoretical Computer Science 268, 91–105.

Fu, L., 2002. A simulation model for evaluating advanced dial-a-ride paratransit systems. Transportation Research A 36A, 291–307.

Goeddel, D., 1996. Benefits Assessment of Advanced Public Transportation Systems. US Department of Transportation, Washington, DC.

Haghani, A., Banihashemi, M., 2002. Heuristic approaches for solving large-scale bus transit vehicle scheduling problem with route time constraints. Transportation Research A 36A, 309–333.

Hauptmeier, D., Krumke, S.O., Rambau, J., 2000. The online dial-a-ride problem under reasonable load. Lecture Notes in Computer Science 1767, 125–133.

Higgins, L., Laughlin, J.B., Turnbull, K., 2000. Automatic vehicle location and advanced paratransit scheduling at Houston METROLift. In: Proceedings of the 2000 Transportation Research Board Conference, Washington, DC.

Ioachim, I., Desrosiers, J., Dumas, Y., Solomon, M.M., Villeneuve, D., 1995. A request clustering algorithm for door-to-door handicapped transportation. Transportation Science 29, 63–78.

Jaw, J.J., Odoni, A.R., Psaraftis, H.N., 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, A., Madsen, O., Solomon, M., 2002. Partially dynamic vehicle routing models and algorithms. Journal of Operations Research Society 53, 637–646.

Lewis, D., Evans, T., Koffman, D., 1998. Impact of reliability on paratransit demand and operating costs. Transportation Planning and Technology 21, 323–346.

Lipmann, M. et al., 2002. On-line dial-a-ride problems under a restricted information model. Lecture Notes in Computer Science 2461, 674–685.

Lu, Q., Dessouky, M., 2004. An exact algorithm for the multiple vehicle pickup and delivery problem. Transportation Science 38, 503–514.

Lu, Q., Dessouky, M., 2006. A new insertion-based construction heuristic for solving the pickup and delivery problem with hard time windows. European Journal of Operational Research 175, 672–687.

Palmer, K., Dessouky, M., Abdelmaguid, T., 2004. Impact of management practices and advanced technology on demand responsive transit systems. Transportation Research A 38A, 495–509.

Quadrifoglio, L., Dessouky, M., 2007a. Insertion heuristic for scheduling Mobility Allowance Shuttle Transit (MAST) services. Journal of Scheduling 10, 25–40.

Quadrifoglio, L., Dessouky, M., 2007b. Insertion heuristic for scheduling Mobility Allowance Shuttle Transit (MAST) services: sensitivity to service area. In: Computer-Aided Systems in Public Transport, Springer Series: Lecture Notes in Economics and Mathematical Systems 600.

Regan, A.C., Mahmassani, H.S., Jaillet, P., 1996. Dynamic decision making for commercial fleet operations using real-time information. Transportation Research Record 1537, 91–97.

Savelsbergh, M.W.P., Sol, M., 1995. The general pickup and delivery problem. Transportation Science 29, 17–29.

Schweiger, C.L., McGrane, J., 1999. Deployment of technology for paratransit: what are the effects on employees? In: Proceedings of the 1999 Transportation Research Board Conference, Washington, DC.

Shinoda, K., Noda, I., Ohta, M., Kumada, Y., Nakashima, H., 2004. Is dial-a-ride bus reasonable in large scale towns? Evaluation of usability of dial-a-ride systems by simulation. Lecture Notes in Computer Science 3012, 105–119.

Stone, J.R., Nalevanko, A.M., Gilbert, G., 1994. Computer dispatch and scheduling for paratransit: an application of advanced public transportation systems. Transportation Quarterly 48, 173–184.

Stone, J.R., Ahmed, T., Nalevanko, A.M., 2000. Internet-based decision support for advanced public transportation systems technology. In: Proceedings of the 2000 Transportation Research Board Conference, Washington, DC.

Toth, P., Vigo, D., 1997. Heuristic algorithms for the handicapped persons transportation problem. Transportation Science 31, 60–71.

Wallace, R.R., 1997. Paratransit customer: modeling elements of satisfaction with service. Transportation Research Record 1571, 59–66.

Wilson, N.H.M., Sussman, J.M., Higonnet, B.T., Goodman, L.A., 1970. Simulation of a computer-aided routing system (CARS). Highway Research Record 318, 66–76.