THE IMPACT OF DIFFERENT ORGANIZATIONAL FORMS OF ROAD PUBLIC TRANSPORT ON DISTANCE COVERED AND ATMOSPHERIC POLLUTION

Marco Diana[*a], Cristina Pronello[b]
[a]Dép. Economie et Sociologie des Transports, INRETS – Institut National de Recherche sur les Transports et leur Sécurité, Arcueil, France
[b]Dip. di Idraulica, Trasporti e Infrastrutture Civili, Politecnico di Torino, Torino, Italy
marco.diana@inrets.fr, cristina.pronello@polito.it

Abstract

The operation of road public transport services that minimize pollutants emissions plays a key role in improving air quality in urban areas, so that considerable financial resources are employed to buy buses with innovative propulsive systems. However municipalities in poorer countries might not have sufficient funds to endow their fleets with such vehicles. In this paper we assess the possibility of lowering distances travelled and emissions by changing the organizational form of the service. We consider a traditional bus service and a demand responsive service using buses or vans, and we identify the best system for every combination of levels of demand and of service quality, as concerns travels and emissions minimisation. The analysis is repeated for different urban patterns and street networks. Results indicate that the use of a demand responsive service can lower distances covered when the demand is quite low but a good quality is required, provided that service requests are known in advance. Furthermore, the utilization of smaller vans with lower emission factors can give substantial benefits concerning atmospheric pollution, with an increase of kilometres travelled that is often negligible.

Keywords: Public transport; Bus lines; Demand-responsive services; Emissions; Atmospheric pollution
Topic Area: F5 Criteria for Sustainability

1. Introduction

Improving the quality of the air in major metropolitan areas is one of the most important actual concerns of public decision makers. The present inadequacy of tools of intervention periodically causes emergency situations, in which concentrations of some pollutants raise beyond the limits. These situations are faced with draconian measures such as traffic blocks, that of course can only be employed for treating the symptoms but that cannot affect the problem itself.

More strategic policies to solve the problem are generally aimed at encouraging the use of public transport instead of private cars, through appropriate transport demand management (TDM) techniques. Of course, beyond the challenge of diverting trips from private to public modes, these actions are effective in reducing pollution only if the public transport system is sufficiently non-polluting. This goal is usually pursued by focusing on the technology of the propulsive system, and big investments are being made in many cities to endow the vehicle fleets with methane-powered or hybrid buses, or even to build new facilities such as tramways or metro lines. It can however be pointed out that there is a risk of simply moving away emissions from exhaust pipes, since in many cases alternative energy sources such as electricity or hydrogen are still obtained from fossil fuels.

[*] Corresponding Author: Tel. +33 1 47 40 72 54 – Fax +33 1 45 47 56 06

Moreover, such technology-driven approach is quite expensive, so that many local realities (small and medium-sized cities, or even larger agglomerations in developing countries) might have insufficient financial resources in order to take these actions. As a consequence of this, we are for example assisting to an increase of air quality problems in many larger cities located in the southern hemisphere, even if their inhabitants paradoxically have lower mobility levels and use more intensively public transport systems, or even bicycles, than people living in European and North American countries.

In this paper the potential effectiveness of a different approach for tackling this problem is investigated. We assess the possibility of lowering distances covered and emissions through a change of the organizational form of the public transport itself, while keeping constant both the vehicles technology and the quality of the service being provided. The main benefit of this would be to decrease the share of emissions due to public transport also in those realities that cannot employ considerable financial resources to implement more technologically advanced projects. An investment in new technologies is usually required to change the organizational form of the public transport service, but this mainly concerns "soft" ITS (Intelligent Transport Systems) equipments whose cost is a fraction of that of a fleet of vehicles with innovative propulsive system.

Our idea then consists in looking if a demand responsive service would compare favourably against traditional fixed-line buses, concerning distances covered and pollutants emissions, for different demand levels and service qualities. The goal is thus to find the combinations of levels of service and of demand that break-even between the two competing organizational forms of public transport, for different idealized urban contexts. It is worth noting that both services can theoretically be exploited using the same vehicles, although we will also assess what happens if we take full advantage of the potentialities of a demand responsive service by using smaller vehicles, as it is actually done in most of systems currently in operation.

To the best of our knowledge, there is no published research specifically dealing with this topic. Diana (2003) has developed a preliminary study to understand how public transport emission levels would be affected if the actual evening bus service in the city of Turin (Italy) were to be partially substituted by a demand responsive service. The results of this study cannot however be easily generalized, since they refer to the specific situation of that city concerning both the service organization and the demand patterns. There have been over the years many other papers that, broadly speaking, have drawn comparisons between fixed and flexible route services, although their focus is in general on economic aspects; see for example Daganzo (1984), or Chang and Yu (1996) for a more recent review. Of relevant interest is also the paper by Dessouky et al. (2003), that shows how it is possible to take into account environmental impacts in the decision making processes concerning fleet systems.

On a methodological point of view, a simulation approach is used to compare the different systems. In the next section we describe it, defining the idealized service area and the characteristics of the travel demand, as well the concurring systems we are going to study. Section 3 states how can we establish equivalences between different public transport systems concerning the level of service, which of course must be kept constant in order to draw meaningful comparisons. In section 4 we present our experimental design and in section 5 the simulation results. Finally we report some concluding remarks and recommendations for correctly interpreting the research findings.

2. The simulation context

2.1. Service areas and demand densities

Distances covered and related atmospheric emissions basically depend on the service quality we would like to provide. It is intuitive to guess that a higher service quality involves more vehicles in operation, in order to decrease wait times. Furthermore, kilometres covered by demand-responsive services are of course sensible to the actual number of customers. There is however another factor that is very important to investigate, moreover because its influence is not so easy to predict. Road network configuration and, broadly speaking, urban shape play an important role in our discussion, and it is important to understand it in order to determine the applicability of our results in different contexts. For example, the statistical distribution of the demand is more likely to have a peak in a monocentric metropolis than in a sparsely populated suburb. Hence in the following we will consider different urban structures, each one with an illustrative demand distribution, in order to shed some light on the range of application of our findings in more operational contexts.

The cases under investigation will be the following three:

- Firstly, we consider a square area of 25 km^2, with a grid street network whose spacing is 0.5 *km*, and in which the demand density is uniformly generated. Thus, each service demand is generated by drawing the origin and destination points from a square uniform [0, 5] distribution. In doing this we assume that trip origins and destinations are statistically independent. Those points are then assigned to the nearest intersection of the grid, in which are then supposed to be located the stops of both the bus and the demand responsive services.
- Next we analyse a circular area of 25 km^2, with a radial street network composed of 18 diametral lines, so that on the outer edge the spacing between two lines is about 0.49 *km*. Each line is divided in 12 segments, so that there are 12 stops on each diameter besides the central point, that is, one stop every nearly 0.47 *km*. Diametral lines have only the central point as common stop. The demand density in this case is supposed to linearly decrease from a maximum in the city centre to 0 on the outer edge. In other words, pickup and delivery points are statistically independent variables whose distance from the centre is drawn from a symmetric triangular [-2.82, +2.82] distribution, and whose orientation is drawn from a uniform [0, 2π] distribution. Then the points are assigned to the nearest stop as in the preceding case.
- Finally we study the same area and demand distribution as described in the preceding point, but in presence of 6 additional equally spaced road rings around the centre. Those rings then connect the stops along the diametral lines that have the same distance from the centre.

All the competing public transport systems cover the totality of the street networks. In all the three cases, the temporal distribution of the demand will be modelled as a Poisson process.

The first situation, that will be indicated with the letter *G* (grid) is perhaps the most studied in researches dealing with the optimal supply of public transport services, since it is mathematically easier. In addiction, it can be seen as a good starting point for more realistic analyses, for example by dividing an urban area in several regions in which the demand is nearly constant and then applying this analysis on each region. The ring-radial network *RR* can represent the classic monocentric city, and is still of interest when urban sprawl processes have not completely reversed territorial dynamics. The intermediate *R* (purely radial) case is more an abstraction, but it has been introduced considering the fact that in many cities the public transport configuration is strongly monocentric, whereas services between peripheral sectors are weaker and more problematic. We have chosen to take to extreme consequences this phenomenon by eliminating any service that does not pass through the centre, in order to study the behaviour of the competing systems. Of course it would not be wise to organize a demand responsive service that can move only along radial lines, so that it could be also interesting to study an hybrid situation between *R* and *RR*, in which rings are used only by demand responsive services.

2.2. Traditional bus services

The bus service is supposed to cover all the streets of the three above networks. For the *G* case, we will have 22 straight lines of 5 *km* length, that become 18 diametral lines of 5.64 *km* when the considered network is *R*. We add to the latter 6 ring lines, of length comprised between 2.95 and 17.72 *km*, to have the *RR* case. In the following we will call this the *FIX* service. All the vehicles move at the same commercial speed *v*, and all the lines are supposed to have the same headway *h*, even if for the *RR* network case it would surely be more efficient to set the ring lines frequency on the basis of the demand density.

Concerning this aspect, we want to point out that we are not attempting to define the optimal bus service in the three considered contexts. For example, it is well known that the optimal configuration of a bus system in the *G* area is not the one we assumed, L-shaped paths across the grid being more efficient (Newell, 1979). Our goal is rather to model the most likely public transport assets that can be observed in real situations, when the urban structure is similar to one of the paradigms we introduced. Those assets are the product of successive historical evolutions and of social groups interaction, more than of a completely rational optimisation approach, and it is on such systems that we are focusing our attention. Readers interested in the optimisation approach for transit systems are referred to the abundant literature in the sector, see for example Ceder and Wilson (1986) for an introductory overview, Kim and Barnhart (1999) and Ceder (2002) for a more updated state of the art review or Baaj and Mahmassani (1995) and Ceder et al. (2002) for applicative studies.

2.3. Demand responsive services

As an alternative to the fixed-route services described in the preceding subsection, we will consider demand responsive systems, operating in the same area, that serve the same demand levels. We will take the organisational form of public transport that is at the opposite of conventional bus lines, namely many-to-many services without predefined paths or schedules that can serve requests between any origin-destination pair without vehicle changes. Among the possible operational specifications of similar services, the one that has been originally proposed by Jaw et al. (1986) is one of the most used, since it considers features that increase the realism of the modelling exercise, and will thus be adopted.

According to this formulation, any customer has to reserve the trip before using the service. He specifies the origin and the destination of the trip, as well as the number of passengers and either the pickup or the delivery time. The trip origin and destination can be any stop of the fixed-route service. The operator fixes (or negotiates) the maximum ride time and the maximum wait time *WT* at the pickup point (for customers that specify the pickup time) or the maximum advance time *AT* at the delivery point (for customers that specify the delivery time), and then schedules the service on the basis of mathematical programming techniques. The scheduling activity basically consists in defining the paths that all the vehicles will have to follow through the given street network, in order to service all the requests in due time. In order to successfully serve a request, the vehicle to which the request is assigned has to visit both its pickup and delivery point within a certain interval, called time window. The maximum ride time for each customer *MT* is the upper bound of the detours that he will have to suffer while onboard, in order to allow the vehicle to serve more than one request at a time.

The lengths of time windows and of the maximum ride time are the two parameters that control the quality of the system. Tightening them ensures that customers will be carried quickly and closer their desired time, but decreases the probabilities of sharing a ride, thus incrementing both the number of needed vehicles and the kilometres to be covered. The maximum ride time is usually set as an increasing function of its direct ride time *DT*, defined to be the travel time if the trip would be made without deviations. We use the following definition for *MT*, where *a* and *b* are two parameters that are specified by the scheduler:

$$MT = a\,DT + b\,. \qquad (1)$$

In our simulation the vehicles are supposed to move at a constant commercial speed *v*, equal to that of the fixed-route service, and can stop and idle waiting for customers at any node of the street network, provided that no customers are already onboard. In the present research we will use the algorithm that has been presented in Diana and Dessouky (2004) in order to schedule our system. We will consider only static services, that is, all the requests come to the dispatch centre before starting the scheduling phase. Given the set of requests, several simulations are run, progressively lowering the number of vehicles in activity, until some request cannot be served without violating some of the problem constraints. Hence we determine the minimum required fleet size, and we deduct from the corresponding schedule the distance globally covered, to be used in our analyses. It is however worth pointing out that we are not determining the minimum distance needed to serve all the requests with this procedure. Our priority is to minimize the fleet, since this is standard practice, but it is generally possible to find solutions that use more vehicles while lowering total runs.

As we mentioned in section 1, one of the advantages of demand responsive services is the possibility of using smaller vehicles, since it is seldom possible to share a ride among a high number of passengers, moreover if time windows are tight and the spatial and temporal demand density not too high. In those cases there is evidently an important benefit concerning fuel consumption and emissions, whereas distances covered are almost the same or slightly increase at worst. Hence, in the following we will run simulations for two kinds of service:

- The first will use the same homogeneous fleet of the *FIX* system (even if the number of vehicles in operations could be different). In this way, the different emission levels are to be exclusively imputed to different organizational forms of the system; we will call this the *basic demand responsive transport service* (*DRTb*).
- The second will use smaller vehicles, namely vans with 8 passenger seats. In this case the corresponding variation in the production of pollutants will be the combination of both the change of the organizational form and of the kind of vehicle; this will be our *smart demand responsive transport service* (*DRTs*).

The two systems are scheduled by the algorithm in exactly the same manner, but the lower capacity limit in *DRTs* could of course induce the need of using more vehicles and driving more kilometres to serve all the requests, at least when there are good rideshare possibilities (high demand and low service quality). In this case there is then another trade-off to be investigated.

To sum up, the experimental design, to be defined in section 4, will compare for a given pattern of demand in a given service area the outcomes of three different services: *FIX*, *DRTb* and *DRTs*. As we previously pointed out, our simulations should however be able to keep a constant level of service across these different systems. This important point is developed in the next section.

3. Comparability of the levels of service among *FIX*, *DRTb* and *DRTs*

In order to draw meaningful comparison among different public transport systems serving the same demand, it is necessary to define a methodology to analytically compare their qualities. Remind that we suppose that the two systems are travelling at the same commercial speed v, which includes also the stops to let the people get in and out the vehicle. We will also disregard other possible sources of noise, such as the comfort onboard the vehicle or any other element that could influence customers' perceptions and opinions regarding the service quality.

3.1. Service quality for FIX, grid network case

Given the abovementioned assumptions, the relative quality of *FIX* basically depends on the headway h between two vehicles of the same line. Recall that all the lines have the same headway. For the grid network case, any possible travel involves the use of no more than two different lines. Considering the example shown in figure 1, the customer could move from east to west, and then from north to south (path *I*), or vice versa (path *II*). It can be seen that there are several other possible paths of equivalent spatial length (8 blocks), but that would involve the use of more than two lines (path *III* and *IV* are two examples). In those cases the time spent onboard the vehicle is the same but at each additional bus transfer the customer will have on average to wait for a time period of $h/2$, making the trip last longer.

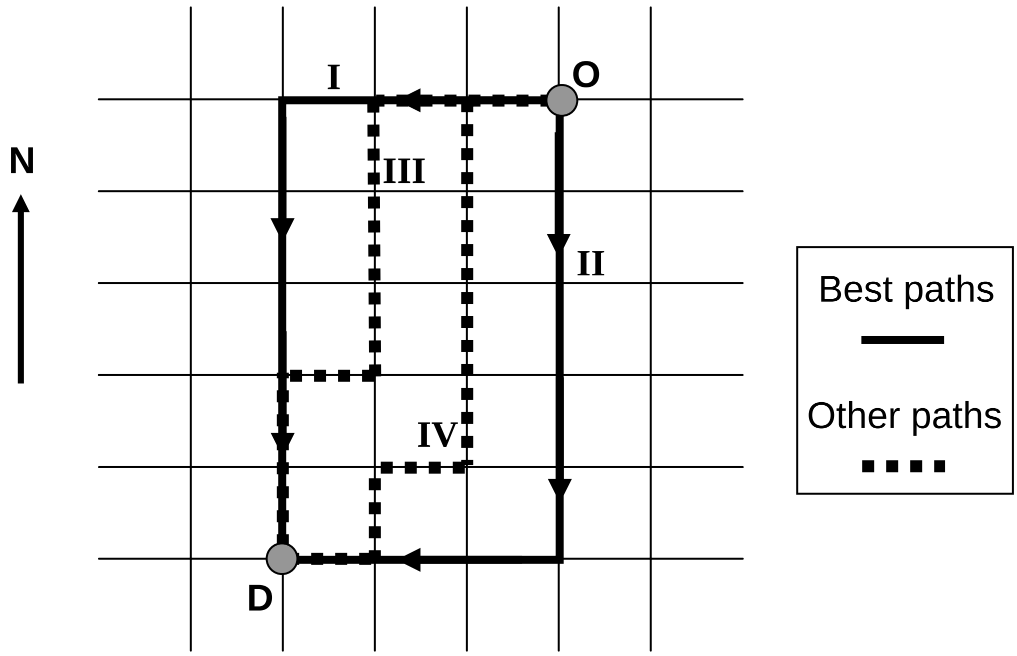


Figure 1. Competing paths of the vehicles for the same origin-destination pair in the grid network case

Let us compare for simplicity the travel time t (including wait times at the pickup and eventually at transfer points) with the travel time t_E associated with the euclidean distance between origin and destination, ideally supposing to cover it at the same speed v and without having to wait for being served. The difference between t and t_E can be looked as an overtime, that is an expression of the level of service for a particular travel, and it is obviously not constant among different trips. Its minimum reached is when the customer has to take only one bus, in which case he will have to wait on average $h/2$:

$$\min (t - t_E) = h/2 \tag{2}$$

These trips with minimum overtime are those with the best possible level of service. On the other hand, the maximum overtime is reached when the line segment linking origin and destination has an orientation of 45° as to the greed network. In this case, the customer will wait on average $h/2$ at the origin point and $h/2$ at the transfer point, and the distance covered will be $\sqrt{2}$ times longer than the euclidean distance between origin and destination:

$$\max (t - t_E) = h/2 + t_E (\sqrt{2} - 1) + h/2 . \tag{3}$$

Earlier results in geometrical probability (Eilon et al., 1971) have shown that the mean distance between two uniformly distributed points in a square is about 0.52 times the side length, so that the above equation can be approximated in the following way:

$$\max (t - t_E) \approx 0.21\, L/v + h \tag{4}$$

where in our case L is 5 kilometres and v is 20 kilometres per hour. This is the longest overtime using *FIX*, corresponding to customers with the worst level of service.

The final passage is to determine the mean level of service between these two case limits. We will assume that the distribution of the overtimes can be approximated by a uniform distribution between the values reported in (2) and (4). Two complications are disregarded when making this assumption. We overlook the fact that the lower limit does not have the term associated with the wait time ($h/2$) at the transfer point. Secondly, there are some "corner effects", due to the fact that travels inside a square longer than L cannot have any orientation. However a closer inspection of the distribution of the distances between two uniformly distributed points in a square, reported in Diana et al. (2004), reveals that the trips longer than L are only about 2.5% of the total, the remainder not being affected by the corner effect. To sum up, we overlook these complications, assuming that the overtime when using *FIX* is uniformly distributed within the interval $[h/2,\ 0.21\, L/v + h]$, so that the expected level of service is given by:

$$E(t - t_E) = 0.11\, L/v + 0.75\, h . \tag{5}$$

3.2. Equivalent service quality for DRTb and DRTs, grid network case

We have now to find the service quality provided by *DRTb* and *DRTs*, that can be comparable to that of the corresponding *FIX* service. This can be done by properly setting the quality requirements followed when scheduling the service, thus conveniently fixing the parameters a and b of equation (1), as well the maximum wait time WT and the advance time AT.

At the outset, we point out that the difference between maximum ride time MT and direct ride time DT can be expressed through equation (1) as a linear function of the direct ride time itself:

$$MT - DT = (a - 1)\, DT + b . \tag{6}$$

This is a nice characteristic, since it is clear that an overtime of, say, 10 minutes will be much more acceptable for a customer whose trip length is 50 minutes than for a customer requesting a 5 minutes travel. However this linear dependency is not present in *FIX*, so that we will get rid of it by setting $a = 1$.

The minimum overtime, min $(t - t_E)$, is reached when a customer travels between two points located on the same street and the vehicle makes no detours. In this case, $t_E = DT$ and customers that have specified the pickup time will have to wait on average $WT/2$ for the bus, so that so that $t = DT + WT/2$, whereas delivery time-specified requests would be serviced on average $AT/2$ earlier than desired, thus leading to $t = DT + AT/2$. In the following we will develop only the former case, since the two are identical, needing only to substitute *WT* with *AT* in the equations. To sum up, we can write

$$\min (t - t_E) = WT/2 \ . \tag{7}$$

The maximum overtime is reached when the line segment linking origin and destination has an orientation of 45° as to the greed network and the bus makes the maximum allowable detours, so that the customer has to stay onboard for *MT*. In this case, $t = MT + WT/2 = DT + b + WT/2$ and, recalling the preceding discussion, we have

$$\max (t - t_E) \approx 0.21\ L/v + b + WT/2 \ . \tag{8}$$

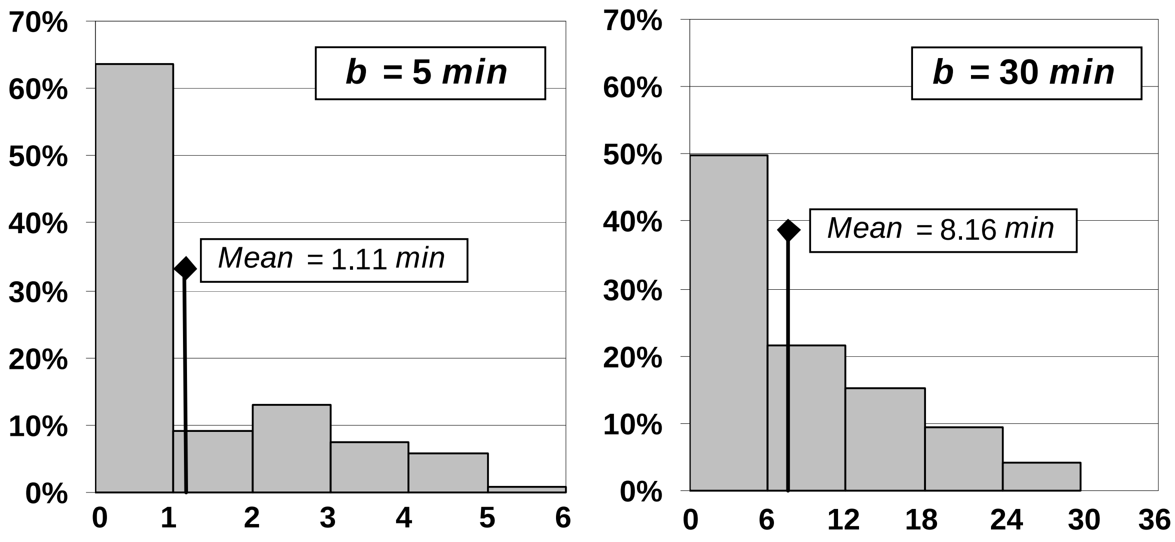
In fact, customers' overtime depends on the orientation of the line segment connecting their origin and their destination in the same manner as described for the *FIX* service, but we do not have to include the time loss at transfer points.

Comparing equations (7) and (8) with the corresponding (2) and (4) allows us to express *WT* and *b* as a function of *h*, that is, to find the *DRTb* or *DRTs* system that ensures the same minimum and maximum level of service of *FIX* with headway *h*:

$$\begin{cases} WT = h \\ b \quad = h/2 \end{cases} . \tag{9}$$

An important difference as regards the *FIX* case however concerns the distribution of the levels of service among different trips. A straightforward interpretation of equations (9) suggests in fact that when we shift from *FIX* to *DRTb* or *DRTs* we transform the expected wait time at the pickup point dependent on headway to an expected wait time due to the time window specification (rewriting the first equation as $WT/2 = h/2$), and the expected wait time at the transfer point to a vehicle detour. However the scheduling algorithm makes use of combinatorial optimisation techniques that will cause a predominance of trips with smaller detours. Equations (9) ensure us that the lower and upper bounds of the level of service are the same in the two cases, but almost surely the mean *DRTb* or *DRTs* level of service is better than that of *FIX*, specified by equation (5). The two would be the same if $b = h/2$ represented the mean detour length in equation (6), and not the maximum possible. Since the distribution of the overtime cannot be controlled from outside but it is a characteristic given by the algorithm itself, the only way to approximate the expected level of service expressed by equation (5) is to alter the value of *b*.

As an indication of the behaviour of the particular algorithm we are going to use, we report in figure 2 the histograms of the distribution of the overtime in two simulations on the ring-radial network of our experimental plan, to be introduced in section 4. They represent the two extreme cases in our study, figure 2a and 2b referring respectively to the simulation with minimum and maximum observed rideshare respectively. An uniform approximation cannot clearly be assumed. Interestingly enough, despite a big difference in rideshare terms, in both cases the number of requests whose detour is the maximum allowable is negligible, or even null. Furthermore, the mean customers' overtime as a percentage of the maximum allowable is not so different (1.11/5 = 22.2% for the simulation with minimum overtime and 8.16/30 = 27.2% for the other one).



(*a*) (*b*)

Figure 2a,b. Distribution of the requests' overtime for the two simulations on the ring-radial network with minimum (*a*) and maximum (*b*) rideshare

The way to proceed depends on the approach we would like to follow when comparing the service quality of different systems. If we want to make sure that the worst levels of service are the same for *FIX* and for *DRTb* or *DRTs*, then we will set the parameters as shown in equations (9). On the other hand, if we are more concerned with the mean level of service, then we will increase *b*, letting some customers of the demand responsive service to suffer an overtime worse than the maximum allowable in the bus system, until we match the mean overtime value expressed in (5). On the basis of further considerations and analyses of the scheduling heuristic that we omit here of briefness, in this research we chose to follow an intermediate approach, increasing the value *b* to heuristically set it equal to the headway *h*. Within the experimental design that will be described in section 4, this resulted in an overall satisfactorily equivalence of the mean level of service provided by the different systems, while keeping the percentage of "sacrificed" requests reasonably low. In fact, considering again the plots of figure 2, it can be seen that the requests in the upper half of the distribution should be less than 20% in both cases. Our choice also allows a remarkable simplification in the presentation of subsequent results, since we can define a *wait state W* as follows:

$$W = h = WT = AT = b\,. \qquad (10)$$

The wait state can synthetically represent the quality level of all our three systems.

3.3. Comparability of service levels with radial and ring-radial networks

The comparative analysis of the levels of service provided by our public transport systems roughly follows the same lines when the considered street network is radial or ring-radial. Algebraic passages will then be omitted, but it is possible to show that the minimum overtime is still given by equations (2) and (7) for the *FIX* and *DRTb* or *DRTs* cases respectively, whereas the definitions of the maximum overtime are different, since they depend on the network geometry and on the distribution of the demand. However the best trip using *FIX* involves no more than two lines (one transfer) for any origin/destination pair also for the radial and the ring-radial network. This makes still possible to "substitute" the wait time at the transfer point into the maximum detour for *DRTb* or *DRTs* customers, and then to define the wait state *W* as given by equation (10).

4. Experimental design

4.1. Experimental context

Our goal is to quantify the domain of each organizational form of public transport service, that is to identify, for each possible combination of factors, the system among *FIX*, *DRTb* and *DRTs* that outperforms the other two concerning distances covered and pollutant emissions. The computation of the distances covered by *FIX* is straightforward, given the headway (that is, in the light of the preceding discussion, the wait time *W*). For the demand responsive systems, it is necessary to schedule the trips and the results depend on the particular sample of requests that is drawn from the demand distribution.

Pollutant emissions have been estimated on the basis of the distances covered following the standard European methodology (European Commission, 1999), not considering cold start emissions since they seem not to be so relevant within our framework. The core of the procedure consists in multiplying the distances by appropriate emission factors in order to obtain the mass of different pollutants, so that, mathematically speaking, emissions are a linear transformation of travelled distances. It can hence be seen that pollutant emissions are to be considered derived result, the focus of the statistical analysis primarily being on the kilometres travelled.

4.2. Design specification

From the previous discussion, we conclude that distances travelled depend on the following factors:

- The kind of urban service area (*G*, *R* or *RR*) and the related distribution of the demand;
- The demand level *D*, that can be seen as the mean of the above mentioned distribution;
- The service quality, synthetically represented for all the systems by the wait state *W*;
- The organizational form of public transport service (*FIX*, *DRTb* or *DRTs*).

When the fourth factor is set to *FIX*, the response is deterministic and does not depend on the demand intensity, whereas in the other two cases it is stochastic. Hence we will design an experimental plan to study the variation of the distances travelled when using *DRTb* and *DRTs*, and then we will compare the results of this statistical analysis with the kilometres travelled by *FIX* at the corresponding combinations of factors. We chose to separately study the three urban areas, each time carrying out an experimental plan with three factors under control (demand, wait time and kind of demand responsive service). Given the preliminary nature of our study, we set up a full factorial design with two levels for each considered factor and two complete replications. It follows that for each kind of urban context we run $2 \cdot 2^3 = 16$ simulations. The levels of the factors under control were set as follows, recalling that when using 2^k factorial designs it is wise to consider the widest possible ranges:

- Concerning the wait state, the low level was set to W = 5 minutes, since it is difficult to be even more punctual in standard urban traffic conditions. The high level was fixed to W = 30 minutes, that is a typical requirement for low-quality paratransit systems, or for door-to-door services in which the customer can wait at home.
- As regards the mean demand, the low level was set to D = 2 requests per minute and the high level was set to D = 50 requests per minute. Within our idealized territorial contexts, the low level corresponds to 4.8 requests per hour per square kilometre. This value is typically encountered in many paratransit services operating either in sparsely populated rural areas or in urban areas to serve specific social groups. The high level is roughly the maximum allowable for a *FIX* system working with the maximum headway of 30 minutes, before reaching the vehicle capacity limit, set to 80 people. In fact, the most heavily loaded system is *R*, since there is the minimum number of lines (18) and only one transfer point. Assuming that 90% of persons uses two lines to accomplish the trip, it follows that there are 5700 trips per hour to serve when we have 50 requests per minute, that is about 316 people that travel on each line in an hour. The vehicle cycle time, given our commercial speed v, is 30 minutes. With a 30 minutes headway, each bus will thus have to carry about 79 people at maximum, assuming that all the passengers are onboard when the centre is crossed.
- Finally, the factor that controls the kind of demand responsive service was transformed into a quantitative variable, considering the number of passenger seats in a vehicle, i.e. the vehicle capacity C. *DRTs* was then supposed to be operated by vans with C = 8 seats, whereas *DRTb* uses larger buses with C = 20 seats. Again, these settings correspond to standard practice in most of existing demand responsive public transport services.

4.3. Length of the simulations

The simulations were run for 2 hours, so that the number of scheduled requests was 240 for smaller problems and 6000 for larger ones. This detail is irrelevant for the *FIX* service but needs to be carefully considered for *DRTb* and *DRTs*. In fact, too short simulation periods risk causing an underestimate of the rideshare possibilities. As a rule of thumb, the vehicle leg, that is the mean travel time between two service nodes, should be at least of one order of magnitude smaller than the simulation length. In our case the more critic situation is encountered considering the G network, since the requests are uniformly distributed and hence we expect longer legs. Conservatively assuming the absence of rideshare, reasoning along the same lines as shown in section 3.1 we can conclude that the expected leg length between two nodes t_N is for the G network

$$t_N = \frac{\sqrt{2}+1}{2} 0.52\, L/v \approx 9.4\, min\,. \tag{11}$$

In the preceding equation the side length of the square is L = 5 *km* and the vehicle speed is v = 20 *km/h*. A simulation of 120 minutes seems thus to give us sufficient margins.

On the other hand, too long periods unnecessarily inflate the number of requests to be scheduled, so that the algorithm might encounter computational difficulties, without giving more reliable results. The algorithm being used was a simple minimum cost-insertion heuristic, described in Diana and Dessouky (2004) as "algorithm 1", whose computational complexity is of $O(n^2)$, if n is the dimension of the instance (in our case, the number of requests). The computational time to schedule 6000 requests was about 8 minutes on a personal computer equipped with a 2.66 *GHz* Pentium® 4 CPU.

5. Results

5.1. Distances travelled

The full 2^3 factorial design we described in section 4.2 allowed us to estimate regression models for the expected number of travelled kilometres y when varying the demand D, the wait state W and the vehicle capacity C. In the latter case we are of course not interested in considering the variation of the response between the two levels of 8 and 20 seats, corresponding to *DRTs* and *DRTb* respectively.

One statistical difficulty we had to address is that the variance of the response is not constant for different factor levels; in particular it increases when we decrease the wait state and we increase the demand density. This phenomenon can be explained considering the scheduling mechanism. Decreasing the wait state lowers ridesharing probabilities, so that the length of the vehicle tours will be more influenced by the variability of the distance between two random points. On the other hand, increasing the demand makes rideshare opportunities more dependent on the two other factors W and C. In other words, if there are few requests, then rideshare is quite unlikely in any case, irrespective of the allowed wait state and vehicle capacity, whereas for high demand levels rideshare and thus travelled kilometres depend on both the wait state and the vehicle capacity. The use of a balanced design, with the same number of replicates for all the experimental points, should alleviate the problem, but we looked for a transformation of the response variable in order to stabilize its variance by using the Box-Cox method. It turned out that the best transformation for the G and R experiments is $y^* = \log(y)$ (transformation parameter $\lambda = 0$), whereas for RR it is better to assume $y^* = \sqrt{y}$ (transformation parameter $\lambda = 0.5$).

We initially considered all the main effects and the two-factors interactions, so that using the above-introduced notation our regression model can be written down as follows:

$$y^* = \beta_0 + \beta_D D + \beta_W W + \beta_C C + \beta_{DW} DW + \beta_{WC} WC + \beta_{DC} DC + \varepsilon \qquad (12)$$

The analysis of variance was used to find the level of significance of each effect. The main effects were always included in the predictive model, regardless their P-value, whereas interaction terms were considered only if their effect was significant at a 5% level. We report in table 1 the value of the coefficients of the predictive models fitted for the G, R and RR cases, and in brackets the P-values of the corresponding term. In all the three cases the adjusted R^2 statistic was above 0.99, but the residual error due to the model lack of fit was found to be significant for the G case. We decided then to look for the presence of significant quadratic effects, and for this we augmented the full factorial experiment to a central composite design, adding the results of 4 more simulations. Since the capacity factor is fixed at the two levels $C = 8$ and $C = 20$, we run 2 additional simulations at each of these levels, setting D = 25 requests per minute and $W = 15$ minutes. However the new analysis of variance did not find significant any quadratic effect, so that those results are not presented here. Deeper statistical investigations could be performed in order to find a model that better fits G results, but we believe that those obtainable from the model reported in the first row of table 1 are in any case fairly reliable for the purposes of our introductory analysis, moreover if we compare them with the outcomes of R and RR.

As expected, the demand and the wait state are highly significant to explain the kilometres travelled by the system, whereas the capacity, that is to say the kind of demand responsive service, is not significant at a 5% level in the G and R networks, although the predicted sign of the coefficient β_C is correct. It seems thus that the use of larger vehicles cannot statistically be proven an efficient mean of increasing the rideshare in two out of three cases, at least within the range of service qualities and demand levels we considered. This finding is not a surprise, since it is well known that the capacity constraint of the underlying combinatorial optimisation problem is weaker than the routing and scheduling constraints. In other words, the street network configuration and moreover the time windows are the stronger limits to the rideshare increase. An analysis across the three different street configurations can however shed some light concerning this point: the *C* factor is totally irrelevant for the *R* network, that is the most inefficient one for a demand responsive service, as pointed out in section 2.1. On the contrary, *DRTb* and *DRTs* work well in the *RR* network, and the weakening of the corresponding routing constraints allows for higher rideshares when the wait time is sufficiently loose. In those conditions, the capacity constraint could effectively be active in limiting the solutions space of the combinatorial problem, as the statistical analysis shows us.

Table 1

Predictive regression models for the grid, radial and ring-radial network cases, P-values in brackets

Network	Response transform	β_0	β_D	β_W	β_C	β_{DW}	β_{WC}	β_{DC}
G	log (*y*)	6.992	5.101E-2 (<0.1%)	-1.962E-2 (<0.1%)	-8.45E-3 (7.0%)	-2.1E-4 (3.0%)	-	-
R	log (*y*)	6.595	5.678E-2 (<0.1%)	-7.598E-3 (<0.1%)	-4.63E-4 (44.7%)	-1.87E-4 (<0.1%)	-	-
RR	$\sqrt{y}$	4.569E-2	-6.86E-4 (<0.1%)	2.32E-4 (<0.1%)	-1.1E-4 (2.3%)	-2.619E-6 (<0.1%)	2.105E-6 (0.2%)	5.678E-6 (1.3%)

Concerning two-factors interactions, D*W is found to be significant in all the three cases. This can be explained considering that an increase in the number of requests to serve has a decreasing effect on the distances covered if the time windows are wider (the sign of the coefficient β_{DW} is negative), since the greater number of requests can partially be accommodated by a rideshare increase.

We report in figures 3, 4 and 5 the contour plots of the regression models for the three urban areas, for both the considered demand responsive systems. We superimposed to those plots vertical dashed lines that mark the headway of the *FIX* system when the kilometres covered are those indicated by the contour plots. It is thus possible to graphically find, for each of the considered six situations, the domain of each public transport system, that is to say the organizational form we should use in order to minimise distances covered under every combination of factors. The headway *h* for every dashed line was computed on the basis of the total length of the network L_N, the distance covered by the bus fleet D_{BUS} and the simulation length *SL*:

$$h = 2L_N\, SL / D_{BUS} = W\,. \qquad (13)$$

Several comments are possible on the basis of these plots. We preliminarily recall that the statistical analysis indicated that the graphically detectable differences between figures 3a and 3b, and 4a and 4b respectively are not significant. However we decided the same to report both plots also for those urban patterns, since in any case those small differences are not counterintuitive. In fact, the contour lines of *DRTs* are slightly shifted towards the bottom compared to the corresponding ones of *DRTb*, due to the tightened capacity constraint. This makes the domain of *DRTs* slightly smaller than that of *DRTb*, and in figure 5 it can be seen that this is particularly true when the wait time increases, as we would logically expect.

Concerning the slope of the contour plots, it is interesting to see that it is greater for the *G* network, so that the response in this case is more sensitive to the variation of the wait state. The boundary line between the two domains shows that demand responsive systems are to be preferred when quality requirements are high and service demand low. Moreover, demand responsive systems seem to be much more competitive than traditional bus services in the case of ring-radial network, where bus services seem to be justified only for very high levels of patronage. This conclusion is valid at least when their headway is not scheduled using a global optimisation approach, as it is the case in most real systems.

5.2. Emissions of pollutants

As it has been mentioned in section 4.1, we will adopt the standard European methodology, based on the COPERT model, in order to assess emission levels for the given distances travelled, considering only hot engine emissions. The pollutants that we quantified are the following five: carbon dioxide (CO_2), carbon monoxide (CO), oxides of nitrogen (NO_x), volatile organic compounds (VOC) and particulate matter (PM). For each of these, the first step is to compute the emission factors, i.e. the grams of pollutant emitted per kilometre of travelled distance. This is a function of the mean speed v and of the kind of vehicle. Concerning the latter point, we would like to avoid introducing a new source of variability, so we will assume a homogeneous fleet of vehicles of EURO III emission class for all the three public transport system. Unfortunately, the methodology embeds all the kind of buses in only one class. For this, it is not possible to distinguish among 20-seat vehicles and the larger buses commonly being used to provide a *FIX*-like service. Hence, we are forced to consider only two kinds of vehicles: EURO III cars for *DRTs* and EURO III buses for the *DRTb* and *FIX*. We report the emission factors for the considered pollutants in table 2.

Table 2

Emission factors for the three public transport systems [*g*/*km*]

System	CO_2	CO	NO_x	VOC	PM
FIX and *DRTb* (a)	1139.31	6.592	7.356	1.729	0.1682
DRTs (b)	215.44	0.471	0.482	0.068	0.0379
ratio (a/b)	5.29	13.99	15.24	27.51	4.43

The last line of the table reports the ratio of the emission factors of each specific pollutant, and is of capital relevance in our analysis. Recalling that we can see the masses of pollutants being emitted simply as a linear transformation of the kilometres covered that we found in the preceding subsection, we can conclude that the comparative analysis we carried out between *FIX* and *DRTb* is still valid when analysing the emissions of the two systems. Hence, figures 3a, 4a and 5a represent the domains of these systems also concerning polluting emission, at least as far as the adopted methodology for computing emission factors does not allow us to discriminate between smaller and larger buses.

Our conclusions drastically change when we consider using smaller vehicles for operating the demand responsive service. The last line of table 2 shows us for example that *DRTs* emissions of CO_2 are 5.29 times less than that of *FIX* when the covered miles are the same. Figures 3b, 4b and 5b clearly show that the *FIX* system can never minimize emissions of CO, NO_x and VOC, irrespective of the demand intensity or the service quality, within the considered range of variation of these factors. However when considering the grid and the radial street network there still seems to be a region in which *FIX* outperforms *DRTs* concerning emissions of CO_2 and of PM, when distances covered to serve the requests are at least 5.29 and 4.43 times less. This result could be particularly significant, due to the attention paid to CO_2 emissions for the limitation of the greenhouse effect and to health problems that PM in particular causes in urban areas. Figure 6 shows the domains of *FIX* and *DRTs* concerning these two pollutants for networks *G* and *R*.

6. Conclusions

In this paper we studied how the organizational form of a public transport service can affect distances covered and pollutant emissions, for different levels of demand and of service quality and in different urban contexts. The outcome of the research has been the definition of the domain of each kind of service, that is the identification of the best one under every combination of the considered factors. Demand responsive services are more effective in minimizing travelled distances when the demand density is not too high but a good level of service is sought, and they show a better behaviour in a ring-radial network. Concerning pollutant emissions, the possibility of using smaller vehicles gives these services an advantage under a wide range of operating circumstances.

Of course caution should be taken when interpreting these results, given the approximations of the COPERT model and the assumptions underlying our demand responsive service operations, particularly concerning the possibility of knowing all the requests before starting the service operation. An interesting extension of the present research could be to compare a traditional bus service with a dynamic demand responsive system. The scheduling efficiency of the latter would be surely lower than the one we achieved here, so that the corresponding domains of *DRTb* and *DRTs* would be reduced. Finally, the idea of using an overcrowded service on demand that runs 5 times more kilometres than a bus, only to have a modest decrease in PM emissions is of course quite utopian, even if considering an idealized context where environmental impacts are the only concern and we totally disregards other elements such as costs. In other words, our results should not be the only factor to consider. They surely represent an interesting element of analysis, to be embedded in a larger multicriteria decision making process, that is the best way of considering environmental issues in transport planning.

References

Baaj, M.H., Mahmassani, H.S., 1995. Hybrid route generation heuristic algorithm for the design of transit networks, Transportation Research 3C (3) 31-50.

Ceder, A., Wilson, N.H.M, 1986. Bus network design, Transportation Research 20B (4) 331-344.

Ceder, A., Gonzalez, O., Gonzalez, H., 2002. Design of bus routes – Methodology for the Santo Domingo case, Transportation Research Record 1791, Transportation Research Board, National Research Council, Washington D.C., 35-43.

Ceder, A., 2002. Urban transit scheduling: framework, review and examples, Journal of Urban Planning and Development 128 (4) 225-244.

Chang, S.K., Yu, W.-J., 1996. Comparison of subsidized fixed- and flexible-route bus systems, Transportation Research Record 1557, Transportation Research Board, National Research Council, Washington D.C., 15-20.

Daganzo, C.F., 1984. Checkpoint dial-a-ride systems, Transportation Research 18B (4-5) 318-327.

Dessouky, M., Rahimi, M., Weidner, M., 2003. Jointly optimizing cost, service, and environmental performance in demand-responsive transit scheduling, Transportation Research 8D (6) 433-465.

Diana, M., 2003. Potential contributions of demand responsive transit services in reducing public transit pollutants emissions in metropolitan areas: an example for the city of Turin, Proceedings of the 10th World Congress on Intelligent Transport Systems (ITS), Madrid.

Diana, M., Dessouky, M.M., 2004. A new regret insertion heuristic for solving large-scale dial-a-ride problems with time windows, Transportation Research 38B (6) 539-557.

Diana, M., Dessouky, M.M., Xia, N. 2004. A continuous approximation model for the fleet dimensioning of many-to-many demand responsive transit services with time windows, submitted to Transportation Science.

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

European Commission, 1999. Meet – Methodology for calculating transport emissions and energy consumption, Luxembourg, ISBN 92-828-6785-4.

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 20B (3) 243-257.

Kim, D., Barnhart, C., 1999. Transportation service network design: models and algorithms, Proceedings of the 7th International Workshop on Computer-Aided Scheduling of Public Transport. Lecture Notes in Economics and Mathematical Systems 471, Springer, Berlin, 259-283.

Newell, G.F., 1979. Some issues relating to the optimal design of bus routes, Transportation Science 13 (1) 20-35.

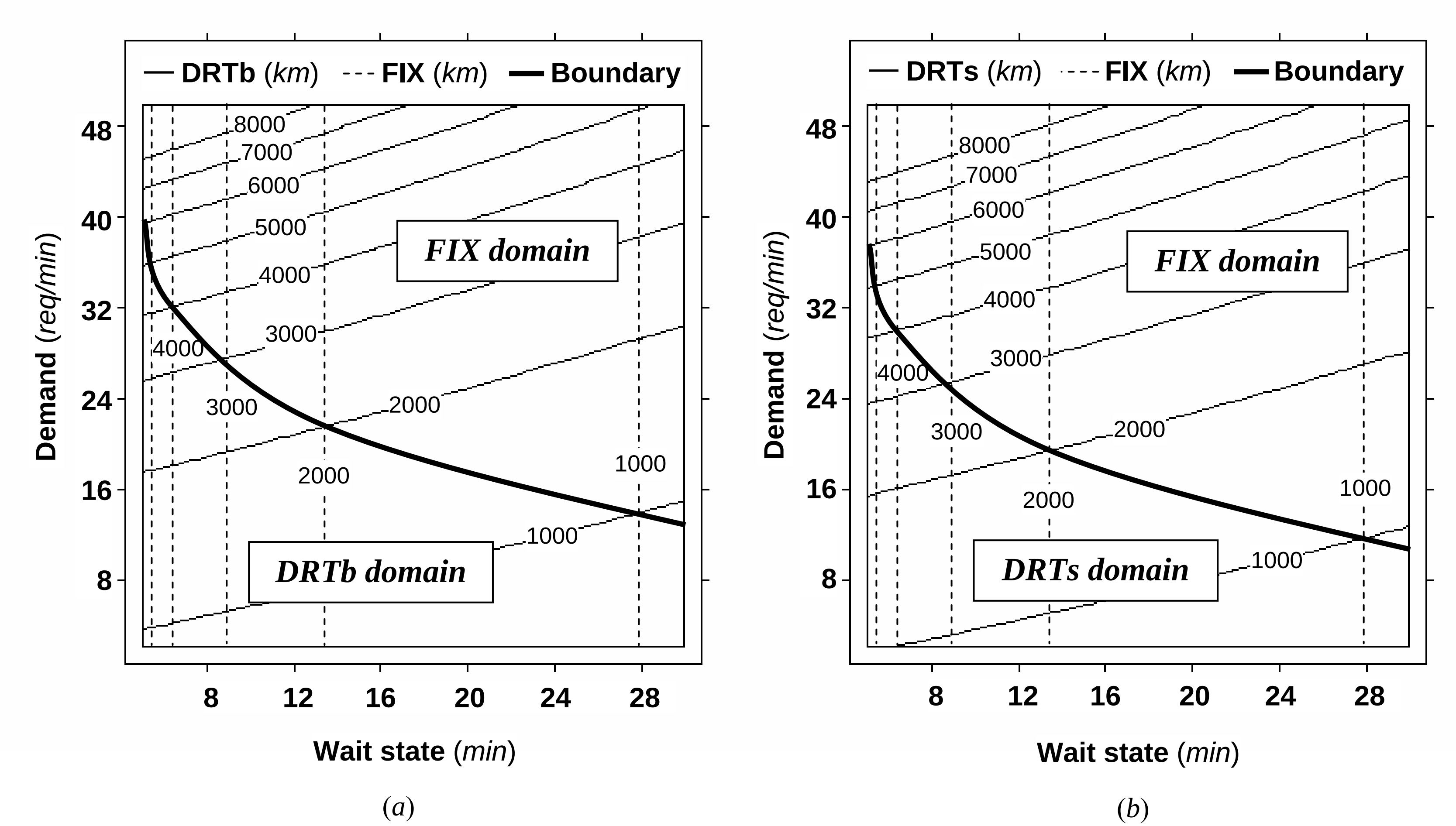
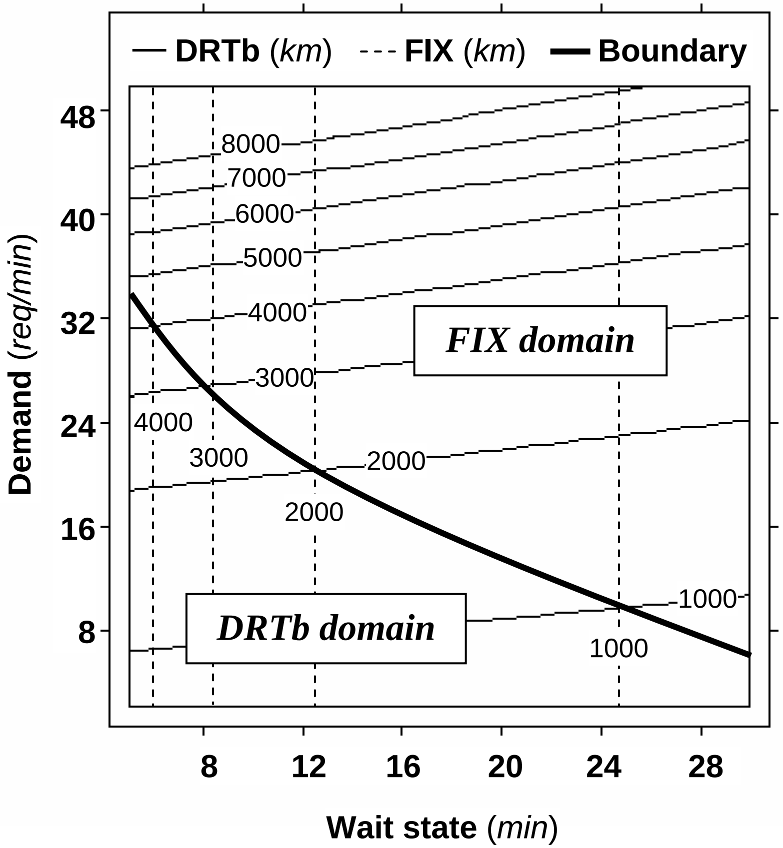
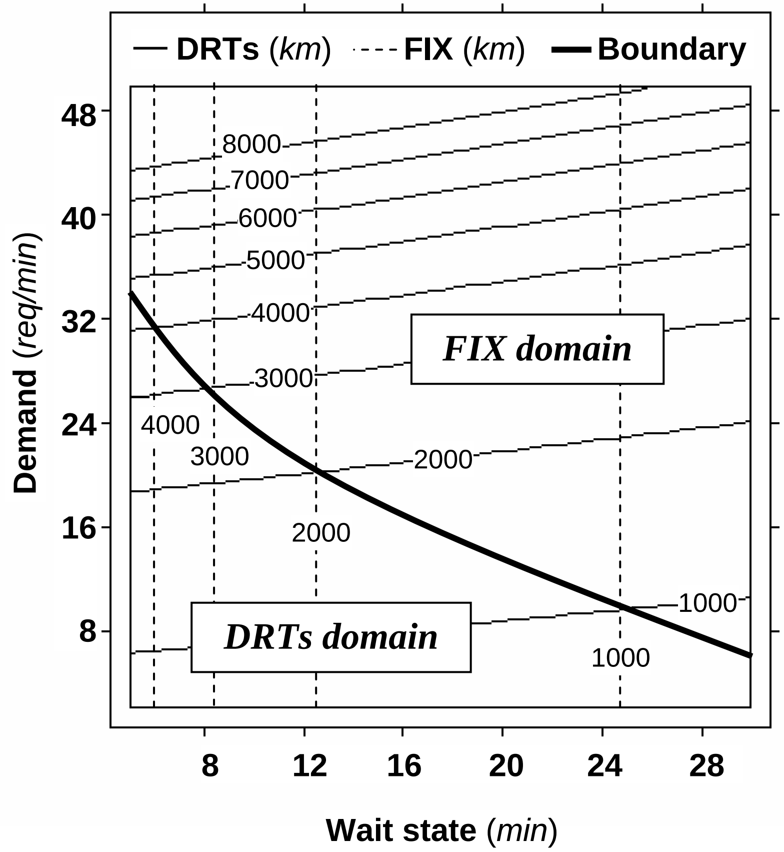


Figure 3a,b. Comparative plots of the kilometres covered by (*a*) *FIX* and *DRTb* and by (*b*) *FIX* and *DRTs*, *G* network case



(a)



(b)

Figure 4a,b. Comparative plots of the kilometres covered by (*a*) *FIX* and *DRTb* and by (*b*) *FIX* and *DRTs*, *R* network case

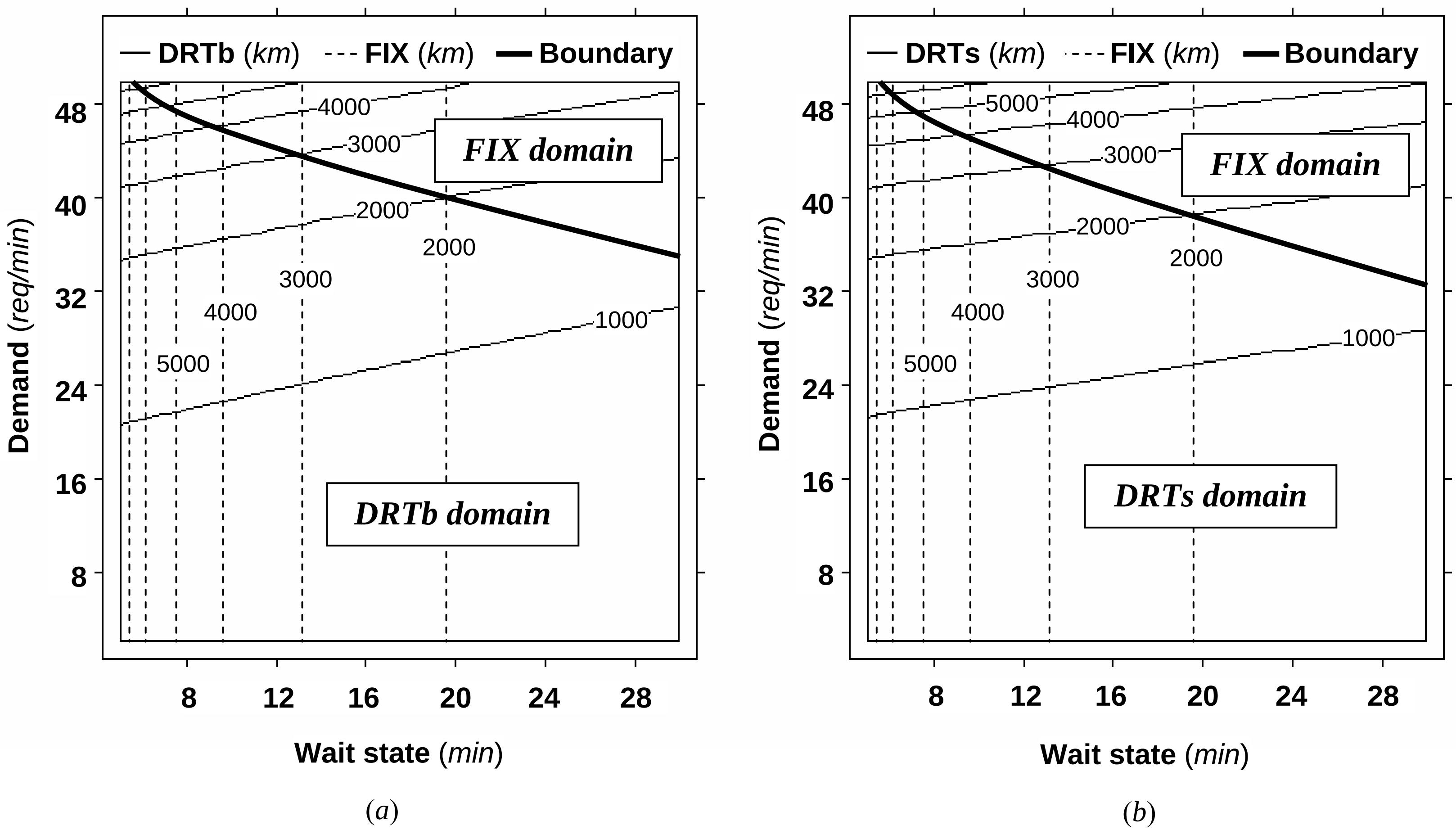


Figure 5a,b. Comparative plots of the kilometres covered by (*a*) *FIX* and *DRTb* and by (*b*) *FIX* and *DRTs*, *RR* network case

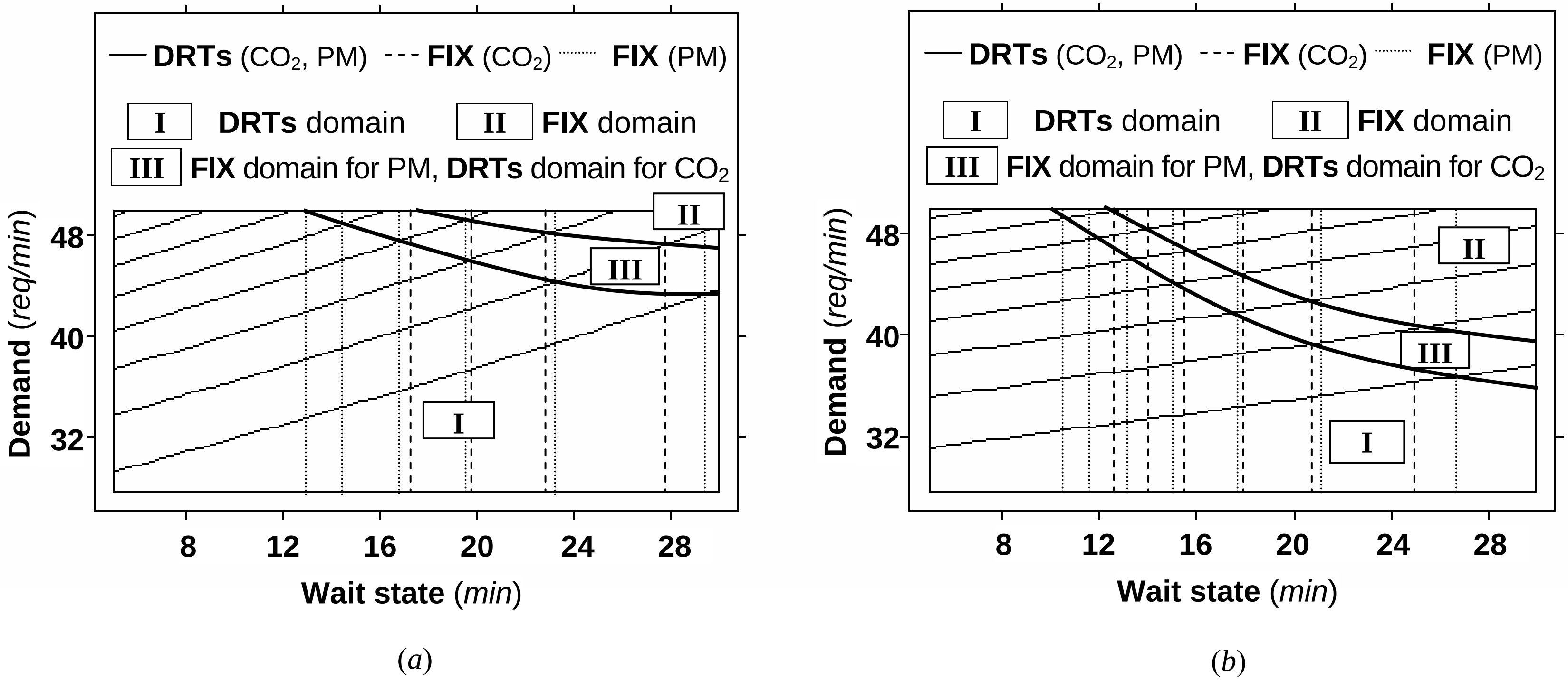


Figure 6a,b. Comparative plots of CO_2 and PM emissions of *FIX* and *DRTs* systems for the *G* (*a*) and the *R* (*b*) network