

## Enhancing flexible transport services with demand-anticipatory insertion heuristics

van Engelen, Matti; Cats, Oded; Post, Henk; Aardal, Karen

**DOI**

[10.1016/j.tre.2017.12.015](https://doi.org/10.1016/j.tre.2017.12.015)

**Publication date**

2018

**Document Version**

Accepted author manuscript

**Published in**

Transportation Research Part E: Logistics and Transportation Review

**Citation (APA)**

van Engelen, M., Cats, O., Post, H., & Aardal, K. (2018). Enhancing flexible transport services with demand-anticipatory insertion heuristics. *Transportation Research Part E: Logistics and Transportation Review*, 110, 110-121. <https://doi.org/10.1016/j.tre.2017.12.015>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

# Enhancing Flexible Transport Services with Demand-Anticipatory Insertion Heuristics

## Abstract

Developments in vehicle automation and the shared economy call for new developments in routing flexible transport services. We propose a new type of insertion algorithm: an online dynamic insertion algorithm with demand forecasts. The performance of this algorithm is tested in a simulation model for a case study network in the Netherlands. When combining the new insertion algorithm with empty vehicle rerouting, 98% of passenger rejections are eliminated and travel and waiting times are reduced by up to 10 and 46% respectively, compared to traditional insertion algorithms. A sensitivity analysis tested performance robustness to variations in operational and demand conditions.

## 1. Introduction

Demand-responsive services have the potential to handle a given demand with fewer resources than line-based public transport, by offering a capacity that is better aligned with the capacity required [17, 18]. Furthermore, demand-responsive services have the additional advantage that they can increase service coverage and accessibility by serving a larger number of stops compared with traditional line-based public transport. However, existing demand-responsive systems are often over three times more expensive than comparable fixed-service systems [2]. This is mainly due to staff costs; demand-responsive services often require a larger number of vehicles with a smaller capacity than conventional public transport services. Staff costs can amount to 70% of total public transport costs in developed countries [7] and public transport operators worldwide are often not profitable and depend heavily on subsidies [12].

Developments in vehicle automation and the shared economy contribute to a growing interest in introducing flexible services. The deployment of automated vehicles will allow operators to dramatically reduce costs and attain greater flexibility in service provision by relaxing many of the existing scheduling constraints. Although these vehicles are not yet available for large-scale deployment, transport operators are already considering how these vehicles could potentially be used to provide efficient and effective public transport in the future. Fully automated vehicles are especially well-suited as an alternative to traditional vehicles for demand-responsive public transport, due to the large amount of personnel required in these systems. In addition, demand-anticipatory capabilities can potentially enable a more efficient and effective deployment of vehicles to travel requests. In this paper we present a new method of rerouting vehicles to customer pickup locations, ideally suited for autonomous vehicles, yet also directly applicable to manually driven vehicles.

The Dial-a-Ride problem consists of designing vehicle routes and schedules for passengers who specify pick-up and drop-off requests between origins and destinations. The Dial-a-Ride Problem (DARP) is an extension of the more intensively studied Vehicle Routing Problem. For an overview of research on the Dial-a-Ride problem, see [6]. Transport is supplied by a certain set of vehicles (possibly with varying sizes) that are based at a set of depots. The most common examples where a DARP is solved are the door-to-door transport of elderly or disabled people (see for example [16, 26, 6, 20]) and taxi services [20, 23]. Another problem that is often tackled as a DARP is the dispatching of ambulance services [9, 24].

The dynamic version of the DARP (known as DDARP) entails that customers' arrival times are not known in advance. Furthermore, the DARP that is considered is stochastic in the following respect: customer requests (their time, origin and destination) are not known in advance, but a probability function of their expected arrival rate is assumed available based on observed historical patterns. [19] offers a recent overview of papers published in this field.

In many cases, ride sharing is allowed in the solution of DDARPs. [1] provide an overview of papers on dynamic ride-sharing. Ride-sharing in solutions to the DDARP is also used by for example [15] and [23] in personal rapid transit and taxis respectively. DDARP formulations vary with respect to service features, such as whether all requests must be served (e.g. [3]), the presence of one-sided time-windows [8], the consideration of fleet size costs [25] or a hybrid formulation that allows deviating from planned routes to accept late requests [4]. Another variant of the DARP considers heterogeneous fleet sizes and fixed costs [10].

The increasing availability of data concerning travel demand and travel patterns presents opportunities for moving beyond demand-responsive services into a proactive mode of operations. Historical passenger data can be used to predict the demand for public transport in the future, in the form of probability distributions. One of the ways that this demand forecast could be used is for dispatching vehicles in anticipation of future demand. We expect such demand-anticipatory operations to result

in shorter passenger waiting times.

In this paper, we propose an insertion algorithm for finding solutions to the Dial-a-Ride Problem which uses demand forecasts for performing rerouting. In particular, we examine the possibilities to incorporate demand forecasts in the routing and timing (hereinafter, routing) of vehicles. We call this new algorithm an online dynamic insertion algorithm with demand forecasts (or, in short, insertion algorithm with demand forecasts). The insertion algorithm works as follows: each time a new passenger request arrives, a choice needs to be made which vehicle will pick up this passenger. This choice can be influenced by using available demand forecasts. For example, a vehicle with only a single empty seat might not be chosen to pick up a passenger if it is expected that more passengers heading in the same direction will arrive shortly at the location of the first passenger. Instead, a vehicle with a higher residual capacity will be chosen to serve the passenger, unlike in existing insertion algorithms.

The proposed insertion algorithm is compared to state-of-the-art algorithms, which do not take passenger demand forecasts into account. Furthermore, the algorithm is combined with empty vehicle rerouting and compared to existing algorithms combined with empty vehicle rerouting. An additional objective of this study is to evaluate the performance of the proposed rerouting strategies for various types of operational and passenger demand conditions. Alternative routing options are evaluated by conducting a series of simulation experiments and analyzing passenger and operator performance indicators for a real-world case study in the Netherlands.

In the next section, the modeling approach is explained, as well as how empty vehicle rerouting and the proposed insertion algorithm are implemented. The simulation tool that is used to determine the effect of the adapted rerouting behavior is described, followed by the details of the case study. Then, the results of the proposed insertion algorithm and empty vehicle rerouting are compared to the results obtained without using the demand forecasts. Finally, we present the main conclusions.

## 2. Method

We solve the Dial-a-Ride problem using an insertion algorithm. In particular, we use an event-based simulation model with a controller that assigns trips to individual vehicles. The network is represented as a graph, with link labels corresponding to the respective travel times between each pair of locations. The model is stop-based, which means that passengers can only be transported between a pre-defined set of locations (termed stops, although these do not necessarily have to correspond to public transport stops). For each public transport stop it is assumed that a demand forecast for how many passenger will request a trip within a given time window is available. Note that it is not required to have a demand forecast for the demand between each pair of stops. The envisaged mode of operations permits the rejection of travel requests if they can not be served within their time windows or the cost is deemed too high by the transport operator.

In the next section, the Dial-a-Ride Problem and the insertion algorithm are described in more detail, followed by the algorithms used for empty vehicle rerouting and the new online dynamic insertion algorithm with demand forecasts. Finally the simulation environment is described in more detail.

### *2.1. Insertion Algorithm as a Solution to the Dial-a-Ride Problem*

In this study we consider the Dynamic Dial-a-Ride Problem (DDARP). In the dynamic variant, some or all of the requests are revealed during the execution of the plan, requiring online decision making. Furthermore, we assume a stochastic process of request generation [22]. This means that a probability function for the expected arrival rates of passengers is assumed available based on archived data. We assume that travel times can be considered deterministic which is reasonable in

areas with low traffic volumes.

Insertion heuristics, first developed by Jaw et al. [13], are used to determine how travel requests are inserted into planned vehicle routes [5, 11]. Demand forecasts have been used to determine slack times in the schedule of a fixed-route shuttle service to allow for small route deviations by Quadrifoglio et al. [21]. In contrast, the insertion algorithm proposed in this study uses the expected demand to determine which vehicle will pick up an incoming travel request.

Whenever a request comes in, a candidate set of vehicles is generated and includes all vehicles that can reasonably serve this request. Then, for each vehicle in the candidate set, all the feasible ways in which this customer can be inserted into the vehicle’s schedule are computed. For each insertion, the value of an objective function is computed. This objective function can be comprised of indicators from both the perspective of the operator and the customer. Furthermore, it is possible to incorporate a penalty for each rejected customer. Finally, the customer is either inserted into the position where the value of the objective function is the smallest or the travel request is rejected (if the value of the objective function for not serving the customer is smaller than the smallest insertion cost). The objective function used in this study consists of three elements: the passenger travel time, the vehicle travel time and the number of rejected requests. The objective function is thus given by:

$$\min_{i \in \mathcal{I}} \left[ \left( \sum_{v \in \mathcal{V}} \alpha \cdot t_v \right) + \left( \sum_{p \in \mathcal{P}} \beta \cdot t_p \right) + \left( \gamma \cdot r \right) \right]. \quad (1)$$

Here  $\alpha, \beta$  and  $\gamma$  are input parameters,  $\mathcal{I}$  is the set of insertion possibilities,  $\mathcal{V}$  the set of vehicles,  $\mathcal{P}$  the set of trips and  $r$  the number of rejected requests.  $t_v$  is the travel time of vehicle  $v$  in minutes and  $t_p$  is the travel time of trip  $p$  in minutes.

## 2.2. Vehicle Rerouting Algorithms

In this section, the online dynamic insertion algorithm with demand forecasts is described in more detail. Furthermore, in many current solution algorithms for the DDARP, an insertion algorithm is used in combination with empty vehicle rerouting. To allow for a sound comparison between the algorithm we propose and current algorithms, this step is also included in our simulations and explained in this section. Furthermore, it also needs to be confirmed that the new insertion algorithms can be combined with empty vehicle rerouting and that they are not mutually exclusive.

### 2.2.1. Online Dynamic Insertion Algorithm with Demand Forecasts

The main goal of this study is to identify to what extent the developed insertion algorithm can be used in combination with demand forecasts to improve the solution to the DARP. The process to include demand forecasts is described here.

Using the insertion algorithm with demand forecasts, vehicles can reroute while traversing a route segment (i.e. en-route between two nodes in the network). An example is shown in Figure 1. Here, the dotted red line between  $d_1$  and  $p_2$  shows the route segment of a vehicle when a request ( $p_5, d_5$ ) comes in. If there is an intersection at point  $i$ , it is allowed for the vehicle to be rerouted such that it heads first to  $p_5$  from  $i$ , before going to  $p_2$ , contrary to algorithms currently used. In Figure 1(b) the dotted line represents the original route of the vehicle.

The demand forecast is used in the following way: when a request  $i$  comes in, first a selection of vehicles that could possibly serve this request is made. Each of the vehicles  $k$  in the selected subset has a certain capacity  $Q^k$  and a certain number of occupied seats  $Q_j^k$  after serving a request at node  $j$  on its route; thus it has  $E^k = Q^k - Q_j^k$  empty seats at that point. Assume that for vehicle  $k$  it takes  $\tau_k$  seconds to serve a request  $i$  after serving request  $j$ . Then, calculate the number of requests

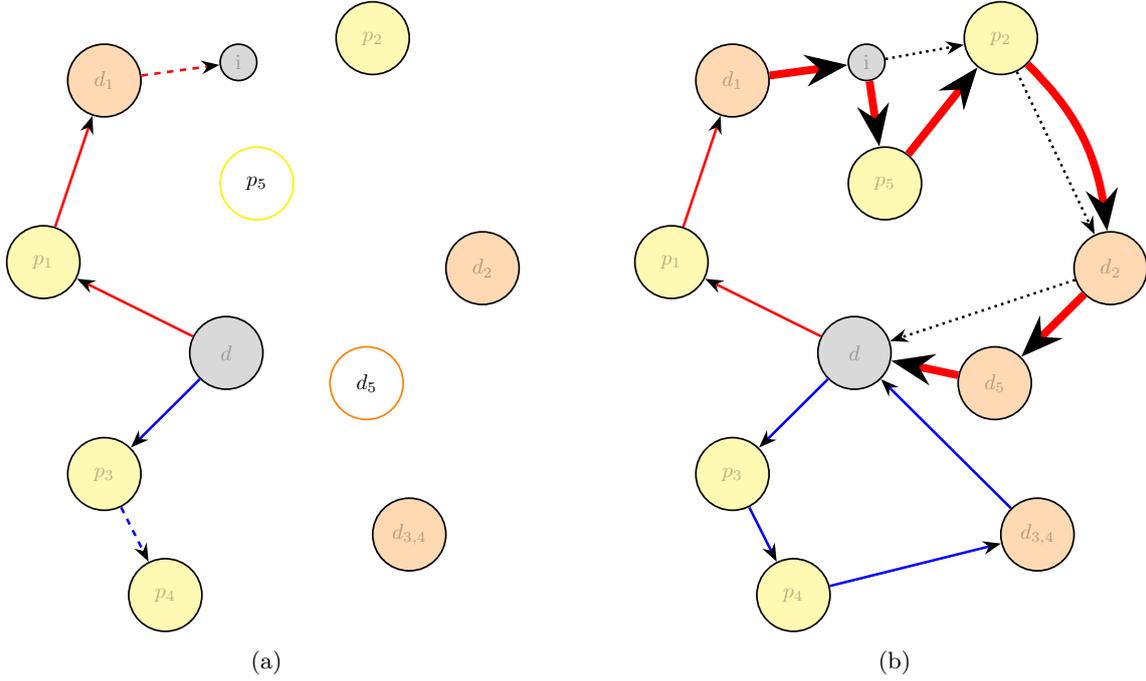


Figure 1: Illustration of the possibility of aborting a route segment. Figure (a) shows vehicles en-route when a new pick-up and drop-off request comes in. Figure (b) shows the adjusted vehicle routes.

that are expected to occur at the node where request  $i$  occurs within  $\tau_k$  seconds of  $i$  occurring, using the demand forecast. Denote this number of requests  $F_i^k$ . Then, if  $E_k < F_i^k$  do not allow the vehicle to serve request  $i$ , unless no other vehicle is available. Based on the risk-taking preference of the transport operator and the trust he places in the demand forecast, the above can be changed to  $E_k < \epsilon F_i^k$ , where  $\epsilon$  is a positive constant. If  $\epsilon < 1$ , indicating a risk-seeking preference, implying that even if a vehicle has less empty seats than the expected number of passengers, it might be allowed to serve the trip. A transport operator could choose  $\epsilon < 1$  if the demand forecast is not very accurate. If  $\epsilon > 1$  it means that a vehicle needs more empty seats than the expected number of passengers, which could be desirable if the transport operator wants to attain a risk-averse vehicle assignment. Note that if  $\epsilon = 0$  the model collapses to the currently used online insertion algorithm.

### 2.2.2. Empty Vehicle Rerouting

For empty vehicle rerouting, the Dynamic Transportation Problem algorithm as it is presented in [14] is used. If  $\mathcal{S}$  is the set of all stops, then for each stop  $i \in \mathcal{S}$  in the network graph, and for each time  $t \in T$ , a target number of inbound vehicles  $\theta_i^t$  is computed, based on the demand at that node. Here  $T = \{0, 1, \dots, t_{max}\}$  is a set of times in the simulation, with a certain interval, where  $t_{max}$  is the total simulation time. The number of inbound vehicles encompasses both the number of empty vehicles that are driving towards stop  $i$  as well as the number of idle vehicles already at  $i$  at a specific time. At each time instant  $t$  the target number of vehicles for node  $i$  can be computed by computing the number of passengers that are expected to arrive at that node within  $x$  seconds, based on the expected value derived from the probability distribution of trip generation. We can then determine, for each stop, the surplus or deficit of vehicles, by determining how many vehicles are currently at each stop, and how many are required at each stop. Then, vehicles are moved from

stops with a surplus of vehicles to stops with a deficit of vehicles as efficiently as possible.

### *2.3. Simulation Flow*

In this section the simulation and its environment are explained. Figure 2 shows a flowchart of the model process. The event-based simulation model progresses by advancing the clock through moving down a chronological event list. The model initialization step involves setting the initial vehicle locations and registering the trip generation events in the event list. If empty vehicle rerouting is part of the simulation, empty vehicle rerouting events are added to the event list as well and empty vehicle rerouting is performed at the model initialization. This ensures that vehicles are located at the nodes in the network where demand is expected, before travel requests are generated. There are three possible events in the simulation, as can be seen in Figure 2:

- In case passengers are picked up or dropped off, routes do not have to be adjusted, as the routes have already been updated to the correct times when the trip has been inserted.
- When a trip is generated, first a time window check is performed, after which all the constraints are checked. If either of these checks fail, the travel request is rejected. Otherwise, the trip is inserted in the most cost-efficient way.
- In the case of an empty vehicle rerouting event, the demand forecasts are used to determine the surplus and deficits at each stop and vehicles are subsequently rerouted as efficiently as possible.

The simulation model is used as a testbed for evaluating the performance of alternative routing strategies under a range of supply and demand settings as detailed in the following section.

## **3. Application**

### *3.1. Case Study*

A real-world case study, the Tata Steel IJmuiden area in the Netherlands, is used to test the different insertion algorithms. This business area is 300 hectares and has a road network that is used by vehicles transporting goods between the different factories, a small amount of private transport by employees, as well as public transport to move personnel between buildings. A map of the case study area is presented in Figure 3. This area was chosen because of its size and isolation. The largest benefits of the proposed insertion algorithm are expected when automated vehicles are used, and while automated vehicles are currently not suitable for driving on roads also used by normal traffic, they can already be used on separate infrastructure (i.e. without interaction with other road users), such as the one available in this business area. Furthermore, given the traffic conditions in this case study area, there are only minor travel time variations and travel times can be considered deterministic.

The network consists of 712 road sections, 36 stops, 327 intersections and 6 depots. For each OD pair, travel times and distances are computed before the simulation starts. The location of stops and depots was decided based on the location of the buildings at Tata Steel IJmuiden, but does not represent their current positions. It is assumed that throughout the morning peak 2000 passengers want to be transported within the case study area.

All simulation runs were performed on Delphi 10 Seattle on an Intel Xeon E5-1620 CPU @3.60GHz with 16GB of RAM using Windows 7.

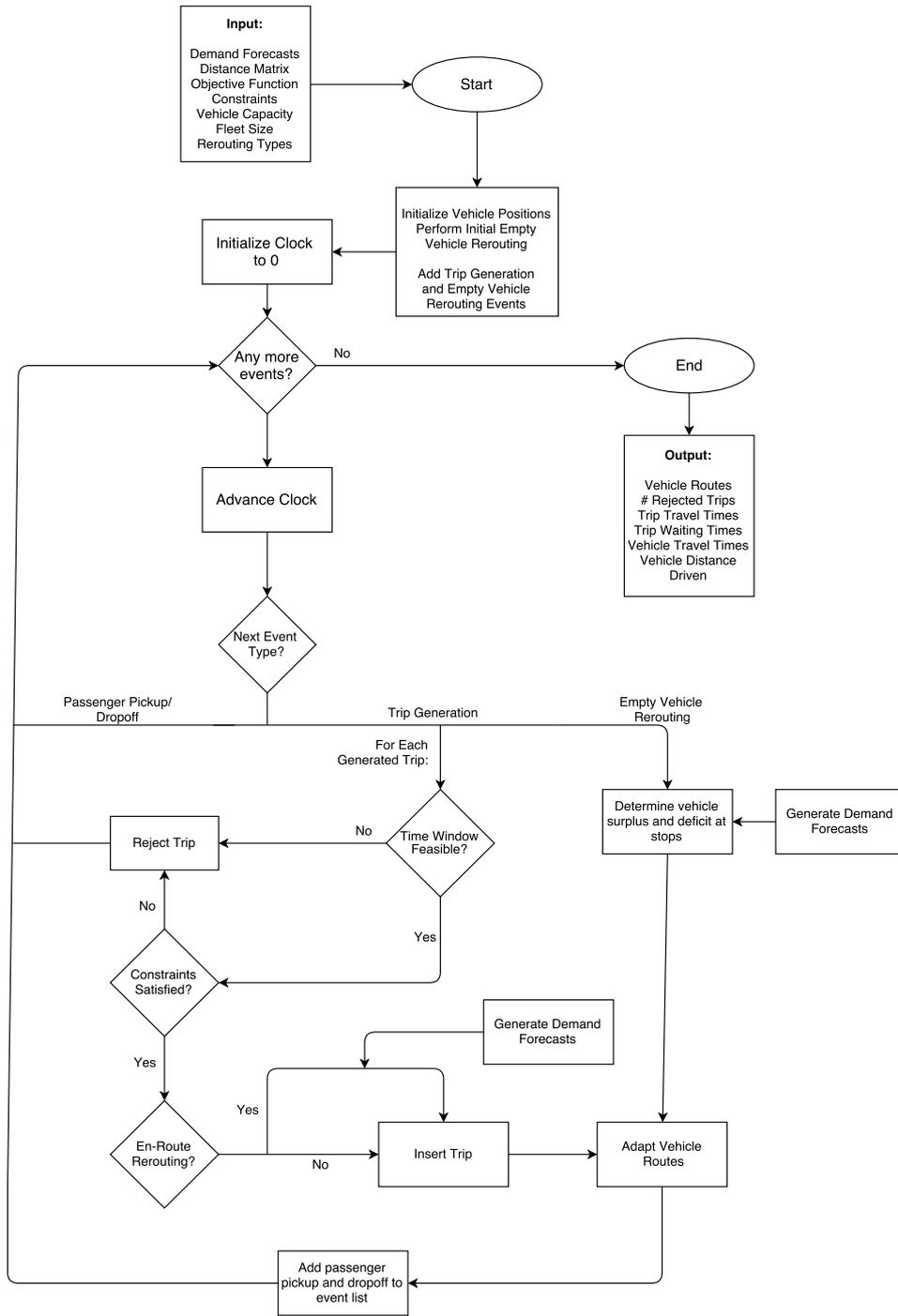


Figure 2: Flowchart of the simulation model

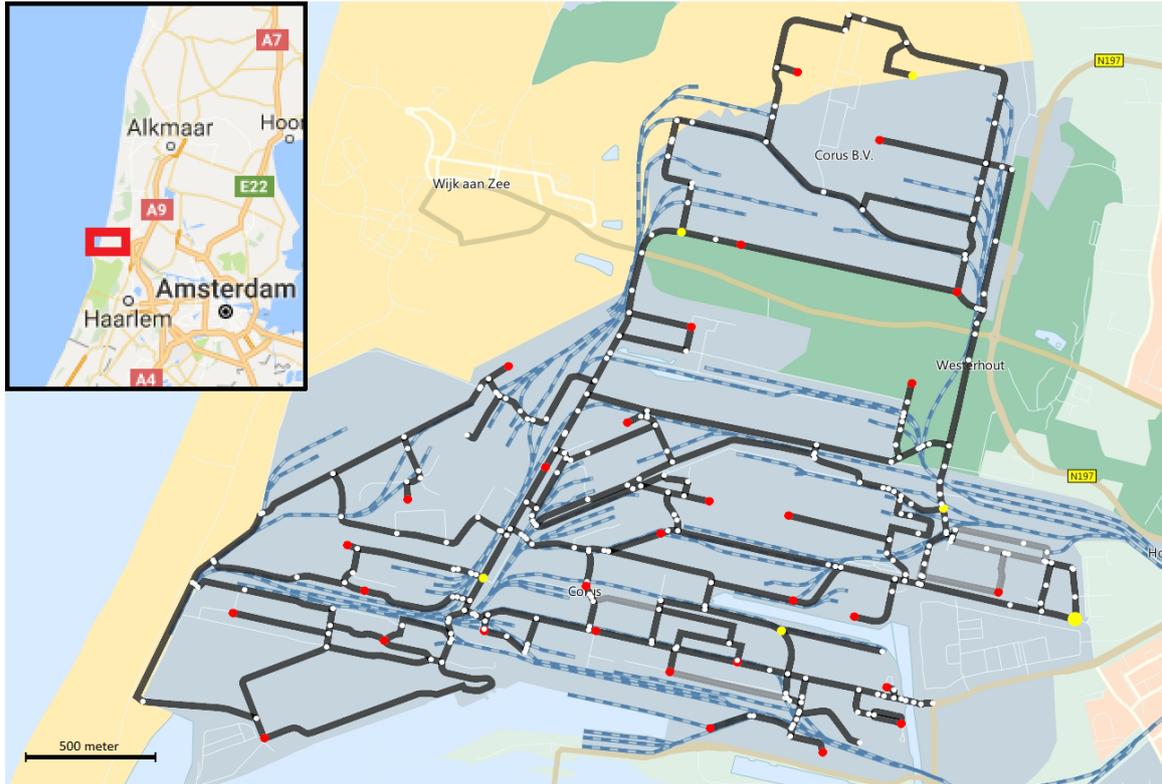


Figure 3: Map of the Tata Steel IJmuiden area in IJmuiden (based on TomTom). The inset (Google Maps) shows the location of the Tata Steel IJmuiden area in relation to Amsterdam. The black lines in the main figure are the roads used in the simulation, gray lines are one-directional roads. The white nodes are intersections, the red nodes are the public transport stops used in the simulation. The yellow nodes are the vehicle depots and the large yellow node is the gate (which is also a depot).

### 3.2. Scenarios

The demand matrix represents a situation where a single location dominates the demand distributions as the primary origin. This location can for example correspond to a gate node through which all external demand enters the case study network, see Figure 3. The rest of the demand is distributed equally over the network. The delivery location of each trip is randomized. The performance of the proposed insertion algorithm is analyzed by comparing the following scenarios:

- **Traditional Insertion** A dynamic insertion algorithm is used, without empty vehicle rerouting. No demand forecasts are used in any rerouting.
- **Traditional Insertion with Empty Vehicle Rerouting** A dynamic insertion algorithm is used in combination with empty vehicle rerouting. Demand forecasts are used to determine where empty vehicles relocate to.
- **Online Dynamic Insertion Algorithm with Demand Forecasts** The online dynamic insertion algorithm with demand forecasts is used without empty vehicle rerouting. Demand forecasts are used to determine which vehicle will pick up each trip. The results of this scenario are compared to the traditional insertion algorithm.
- **Online Dynamic Insertion Algorithm with Demand Forecasts and Empty Vehicle Rerouting** The online dynamic insertion algorithm with demand forecasts is combined with empty vehicle rerouting. Demand forecasts are used both to reroute empty vehicles and to determine which vehicle picks up each trip. By comparing to the scenario where only empty vehicle rerouting is used, we can assess whether the online dynamic insertion algorithm can improve upon currently used insertion algorithms with empty vehicle rerouting.

### 3.3. Model Specifications

The parameters in the objective function were set to  $\alpha = 3$ ,  $\beta = 1$  and  $\gamma = 40$ . The cost for rejected passengers should be high to assure passengers are only rejected in rare cases. The fleet consists of 100 electric vehicles, with a vehicle capacity of 5 passengers. The impact of different fleet compositions on the algorithm performance is investigated as part of the sensitivity analysis. Vehicles are assumed to travel with a constant speed of 40 km/h. The vehicles can each drive a maximum distance of 100 kilometers before their battery has to be charged or they have to be refueled. After 100 kilometers, the vehicles have to visit a depot, where they can be recharged within 20 minutes. Vehicles spend 20 seconds at each stop where they drop off and/or pick up passengers (i.e. dwell time). When a trip request needs to be served by a vehicle, the candidate set of vehicle is created by selecting all vehicles within a driving distance of 2 kilometers.

As the travel times in the case study area are relatively short, a look-ahead time  $x$  of 5 minutes is used for empty vehicle rerouting. The empty vehicle rerouting is performed in the simulation every 2 minutes, as well as 10 minutes before the first trip request arrives. The following constraints are considered as part of the insertion algorithm:

- **Vehicle Capacity**
- **Demand Forecasts** A factor of  $\epsilon = 1$  is used (see Section 2.2.1), meaning a vehicle will not serve a request if the total number of forecasted incoming requests at the request node exceeds its remaining capacity.

- **Passenger Time Window Constraint** This constraint ensures that the passenger that is picked up is dropped off on time at the required destination. The maximum travel time is defined in relation to the direct travel time in both absolute and relative terms:

$$\text{max. total travel time} = \max(\text{direct travel time} + 5 \text{ minutes}, 1.5 \times \text{direct travel time}).$$

- **Vehicle Battery** The final constraint ensures that vehicles are rerouted to a depot when their battery or fuel tank is almost empty. In all scenarios a vehicle is rerouted to a depot once its battery charge is down to a 10% level.

### 3.4. Key Performance Indicators

The efficiency of the model is measured using several key performance indicators. The key performance indicators reflect both the operator costs and passenger costs:

- **Percentage of Rejected Passengers** The share of passengers that are rejected, out of the total number of generated trips.
- **Average Distance Driven per Vehicle** The total distance traveled by all vehicles, divided by the number of vehicles.
- **Average Vehicle Distance per Passenger** The total distance traveled by all vehicles, divided by the number of non-rejected trips.
- **Average Passenger Waiting Time** Average time from trip request to trip pickup time for non-rejected trips.
- **Average Passenger Travel Time** Average time from trip pickup request to trip delivery (i.e. includes passenger waiting time).

## 4. Analysis and Results

In this section the main results of this research are presented, along with an analysis of the spatial distribution of detour and waiting times. A sensitivity analysis of some of the input parameters of the model is performed, including various vehicle capacity and fleet size compositions and variations in the percentage of trips that originate at the entrance node.

Table 1 shows the average key performance indicators for a set of 10 simulations for each of the scenarios defined in section 3.2. A simulation time of 47 seconds was required for simulating 3 hours of real-time transport for the traditional insertion algorithm. 24% of the passengers are rejected because the passengers cannot be transported within their maximum travel time. When comparing the scenarios that do include some type of demand forecasts, we see that passenger travel times and waiting times decrease significantly for both empty vehicle rerouting and the insertion algorithm with demand forecasts.

There are almost no rejected passengers when empty vehicle rerouting is used. Conversely, using only the online dynamic insertion algorithm with demand forecasts does not help to eliminate the share of rejected passenger but yields a high decrease in travel and waiting times. These gains are attained at the cost of a more modest increase in the average vehicle distance per passenger than empty vehicle rerouting. It can be observed that when combining the online dynamic insertion algorithm with demand forecasts and empty vehicle rerouting, that the advantages and disadvantages are combined as well. This means that the two types of rerouting can be combined without counteracting each other. The number of rejections is not reduced beyond what is achieved with only empty

Table 1: Summary of average key performance indicators for 10 simulation runs.

<b>Empty Vehicle Rerouting</b>	<i>no</i>	<i>yes</i>	<i>no</i>	<i>yes</i>
<b>Insertion with Demand Forecasts</b>	<i>no</i>	<i>no</i>	<i>yes</i>	<i>yes</i>
Simulation Time (s)	47	120 (+154%)	53 (+12%)	133 (+180%)
% of Passengers Rejected	24%	0,5% (-98%)	22% (-5%)	0,5% (-98%)
Average Passenger Travel Time (s)	553	534 (-3%)	528 (-5%)	496 (-10%)
Average passenger Waiting Time (s)	139	117 (-16%)	97 (-30%)	75 (-46%)
Average Vehicle Distance (km)	28	45 (+60%)	31 (+9%)	47 (+66%)
Average Veh. Dist./ Passenger (km)	1,87	2,28 (+22%)	1,99 (+6%)	2,37 (+27%)

vehicle rerouting. The reason for this is that empty vehicle rerouting already eliminates the number of rejected passengers almost entirely. For the passengers that are still rejected it is simply not possible to deliver these passengers within their allowed time windows, regardless of the vehicle routing strategy employed. Insertion with demand forecasts only deals with choosing which vehicle will service which trip request, but does not cause vehicles to arrive at the trip departure location earlier. Although the simulation time goes up, especially for empty vehicle rerouting, this still does not jeopardize real-time applicability of the algorithm. Even when combining the two types of rerouting, the complete 3 hour simulation can still be run in just over 2 minutes.

#### 4.1. Detour and Waiting Time Distribution

It is important to analyze the detour and waiting time distributions throughout the network as service experience is determined not only based on the average conditions but also on service reliability, an important determinant of customer retention. The positive effects of the proposed insertion algorithm are somewhat negated if there is an advantage for trips at a few locations, but a large disadvantage for trips at some other locations.

Figure 4 displays the cumulative detour and waiting time distributions for the case study application. When using the online dynamic insertion algorithm with demand forecasts but no empty vehicle rerouting, the waiting time distribution clearly shifts leftwards, i.e. from longer waiting times to shorter waiting times, whereas detour times become slightly longer. This is to be expected, as it is often the case that a vehicle with a larger residual capacity is sent to pick up a passenger, even if that vehicle is slightly further away. When using only empty vehicle rerouting, waiting times are shortened along with considerably shorter detour times. In particular, many previously medium detour times become significantly shorter. When combining the two, the benefits of both algorithms are again combined: waiting times decrease most, without negatively affecting very long waiting times; detour times remain similar to using only empty vehicle rerouting.

Besides the distribution of the length of waiting and detour times, introducing the proposed insertion algorithm could change the location in the network where most detours take place. It is possible that there are for example a few locations where passengers will have very long waiting or detour times. In order to investigate such potential consequences, we contrast the average waiting and detour times for each stop in the network for the traditional insertion scenario without empty vehicle rerouting and the scenario with both the online dynamic insertion algorithm with demand forecasts and empty vehicle rerouting, depicted in Figure 5. Larger nodes in these figures indicate a higher average detour or waiting time. Noticeable decreases in detour times when using the proposed algorithm are spread evenly across the network, and there is not a single location where average waiting time exceeds 3 minutes. Waiting times are reduced, albeit moderately, in a similar manner, with a slight increase in average waiting time occurring at only two nodes. In conclusion, the results confirm that the

proposed insertion algorithm has no negative effects on the spatial distribution of waiting and detour times for passengers.

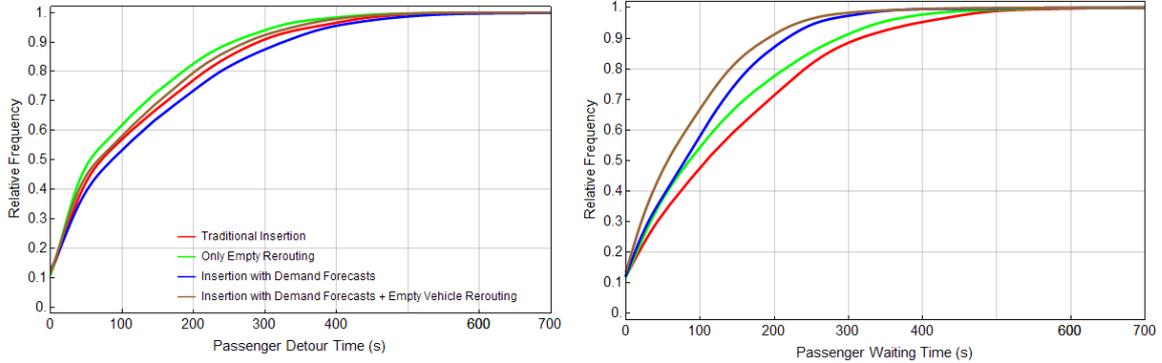


Figure 4: Detour (left) and waiting time (right) distributions.

#### 4.2. Vehicle Capacity and Fleet Size

There is always a trade-off between vehicle capacity and fleet size, the key determinants of service capacity. The results reported so far were analyzed using a fleet size of 100 vehicles with a capacity of 5 passengers each. However, a change in fleet size or vehicle capacity could have an impact on the efficiency of the proposed insertion algorithm. In order to assess their influence, the key performance indicators for several fleet sizes and vehicle capacities are investigated. The total capacity of the system is kept constant (e.g since the standard is 100 vehicles with a capacity of 5 passengers, we use 50 vehicles with a capacity of 10 passengers and similarly for other fleet sizes). Figure 6 presents the key performance indicators for each of these fleet size-vehicle capacity combinations, comparing the traditional insertion algorithm with the proposed algorithm combined with empty vehicle rerouting.

Passenger travel times decrease when more vehicles are available and hence ride-sharing occurs less frequently. As can be expected, waiting times slightly increase when using fewer vehicles, although only slightly if empty vehicle rerouting and the insertion algorithm with demand forecasts are used. In the case of vehicles with a capacity of three passengers, waiting times increase tremendously because there are very few opportunities for ride sharing. Only passengers with very short direct travel times are able to be served at all. Applying the proposed insertion algorithm results in large waiting time savings (i.e. a reduction of 75% in case of a vehicle capacity of three passengers) due to increased ride-sharing and a greater likelihood that a vehicle is readily available close to where travel requests are generated. When reducing the number of vehicles and increasing the vehicle capacity, travel times increase, as the larger vehicles allow for accommodating more ride-sharing trips. Rejections steadily rise when using fewer but larger vehicles.

In general, larger fleets of smaller vehicles are superior to smaller fleets of large vehicles. The former result in shorter waiting times, fewer detours and fewer passenger rejections. This can be expected since large fleets of low-capacity vehicles are better suited for flexible on-demand services in systems with low to medium demand levels from the passenger's point of view. However, a larger fleet with smaller vehicle capacities performs worse than the original vehicle fleet, when considering the total vehicle distance traveled, which is an important performance indicator for the operator as well as any governmental agencies, due to the induced transport externalities (i.e. congestion, emissions). Overall, changes in fleet size and vehicle capacity have a clear impact on the effectiveness of the

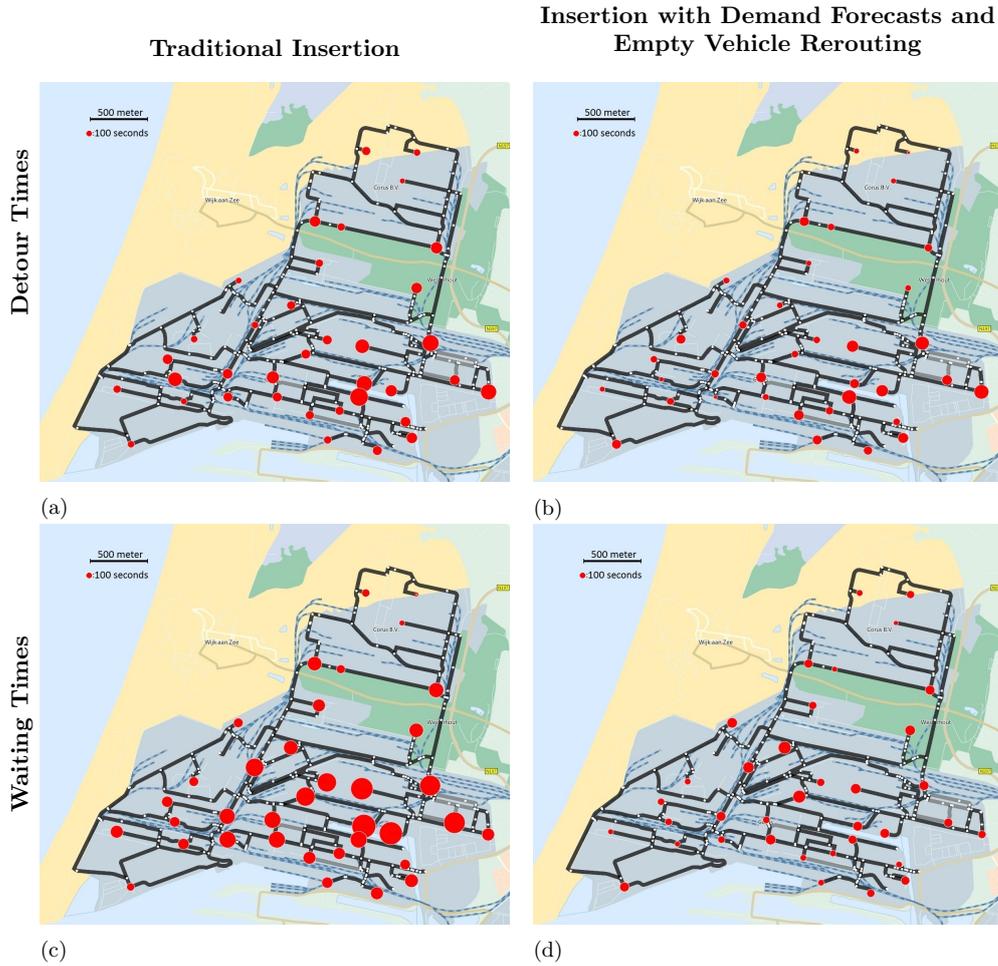


Figure 5: Waiting and detour time spatial distribution for traditional insertion and the online dynamic insertion with demand forecasts combined with empty vehicle rerouting.

insertion algorithm. The percentage of rejected passengers is similar for a vehicle capacity of five or less, but then steadily rises as larger and larger vehicles are used, and the benefit of using demand forecasts in routing vehicles all but disappears when 50 vehicles with a capacity of ten passengers are used. Average passenger travel times increase with an increased vehicle size and reduced fleet size, whereas passenger waiting times remain at the same level. The benefits in travel times attained by the new insertion algorithm decrease with larger vehicles, but the benefits in waiting time stay relatively constant.

As mentioned before, one of the main benefits of using automated vehicles is that more but smaller vehicles can be used due to the removal of personnel costs, allowing to better cater for passenger demand. These results demonstrate that using smaller vehicles is indeed beneficial in reducing the percentage of rejected passengers and passenger travel times, at the cost of increasing the total travel distance. The ideal vehicle size arguably depends on travel distances and demand concentration. The effect of the latter on the performance of the insertion algorithm is investigated in the following

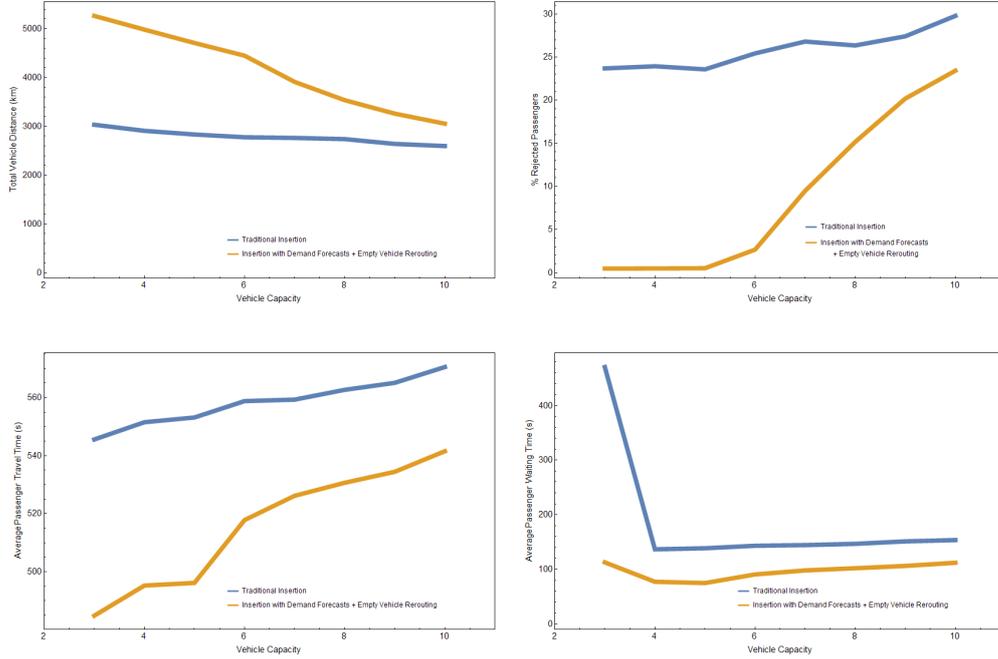


Figure 6: Key performance indicators for variations in fleet size and vehicle capacity, comparing a traditional insertion algorithm with the online dynamic insertion algorithm with demand forecasts.

sub-section.

#### 4.3. Dominating Demand Trip Distribution

To analyze the performance of the insertion algorithm for various spatial distributions of trip requests, the percentage of passengers that originates at the main node was varied. Figure 7 shows the key performance indicators for the variation of this percentage from 0 to 100%. This allows for investigating the consequences of demand patterns ranging from evenly distributed demand to all demand generated in one node, i.e. a 1:N feeder service. The share of unsatisfied travel requests increases exponentially in relation to increased demand concentration (top-right in Figure 7) if no anticipatory capabilities are deployed. This share increases quickly to levels that are likely to be considered unacceptable by service users. A responsive mode of operations imply that vehicles remain idle after dropping-off passengers unless assigned to new trips, and only bounce back when a travel request is generated. The combination of empty vehicle rerouting with the online dynamic insertion algorithm reduces the share of rejected requests to less than one in 200 requests, and this performance is robust to all possible demand concentration scenarios. The waiting time increases to an average of 250 seconds when 100% of the demand originates at one single node, which directly leads to an increase in passenger travel times as well. The increase in waiting time occurs because all vehicles always need to travel back to the entrance, no matter where they dropped off their passengers, due to the strictly unidirectional demand pattern. When a lower percentage of trips originates at the entrance, the vehicles can pick up a passenger close to the drop off location of their previous passenger. However, the more demand originates from one location, the higher the percentage of

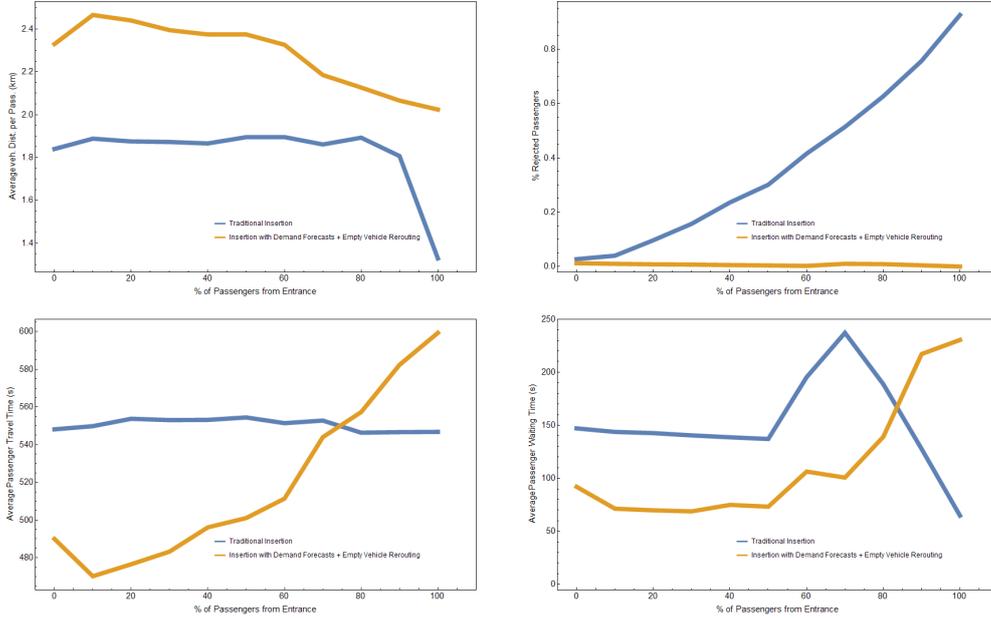


Figure 7: Key performance indicators for variations in the percentage of passengers originating at a central node, comparing a traditional insertion algorithm with the online dynamic insertion algorithm with demand forecasts.

ride sharing is, which in turn leads to a smaller value of the average vehicle distance per passenger. In conclusion, demand concentration has a profound effect on the key performance indicators. While the increase in passenger waiting times could be alleviated to some extent by using a larger fleet, this will increase fixed and variable costs and will lead to an increase in vehicle distance driven. Notwithstanding, considering the number of rejections when demand-anticipatory strategies are not used, there is a strong argument for using the proposed insertion algorithm and empty vehicle rerouting, regardless of the percentage of trips originating at a single location.

## 5. Discussion and Conclusions

### 5.1. Key Findings

Many transport operators worldwide are currently in the exploratory phase of developing their plans for automated vehicles in their operations. Transport operators in many countries are also gathering large amounts of data about the travel behaviour of their customers; however, this data is often not used extensively yet. Furthermore, transport operators are facing increasing competition by new companies offering ride-sharing services. Consequently, transport operators are forced to think about how they too can implement demand-responsive transport, as it often better serves the travel demand and is more suitable for dealing with demand fluctuations.

This study shows that the use of demand forecasts can improve demand-responsive transport significantly and allow it to become anticipatory rather than merely responsive. We propose and demonstrate that demand forecasts can be used to improve the routing of vehicles in the Dial-a-Ride problem in two ways. First, empty vehicles can be rerouted to stops where demand is anticipated.

Secondly, if a new request comes in, the choice of which vehicle should pick up this request could be changed by analyzing the demand forecast at the stop where the request is located.

The effectiveness of these rerouting measures was examined for several different scenarios of the case study network of the Tata Steel IJmuiden area. The key findings are that:

- Empty vehicle rerouting can lead to a dramatic reduction in the number of rejected passengers of up to 98%. Furthermore it can reduce total passenger travel times and waiting times by up to 5 and 16% respectively. The main trade-off is that the total distance driven by all the vehicles increases by up to 60%.
- The online dynamic insertion algorithm with demand forecasts has a less pronounced effect on rejections, reducing them by up to 7%. However, average passenger travel and waiting times can be reduced by up to 6 and 30% respectively, while only requiring an increase of up to 9% in the total distance driven by all vehicles.
- Using both both empty vehicle rerouting and the online dynamic insertion algorithm with demand forecasts results in the summation of their benefits. The number of rejected passengers can be reduced by 98%, while passenger travel and waiting times are down by 11 and 46% respectively. This induces an increase in the vehicle distance driven of up to 66%. However, because of the large reduction in rejected passengers, this translates to only a 25% increase in the vehicle distance driven per passenger.

In conclusion, when demand forecasts are used for empty vehicle rerouting and in the insertion algorithm, almost no passenger demand is left unsatisfied. Furthermore, passenger travel and waiting times can be reduced significantly. This comes at a cost in vehicle distance driven per passenger. However, when automated vehicles are used, this increase in vehicle distance driven is not as significant, since operational costs for vehicle mileage are much lower than when using non-automated vehicles. Costs for externalities are low in non-congested conditions such as the case study area.

A sensitivity analysis of these results shows that the effects of using the proposed insertion algorithm are apparent in all but the most extreme cases of demand distribution, where more than 80% of the transport demand originates at a single location. There are benefits to using the algorithm regardless of fleet composition.

In conclusion, demand forecasts can be used to significantly improve the routing in a Dial-a-Ride problem. Assuming that transport operators do not want to reject too many passengers, as service availability is essential for customer loyalty and retention, empty vehicle rerouting should always be included in algorithms to solve the Dial-a-Ride problem. However, the addition of demand-anticipatory qualities to the insertion algorithm, combined with empty vehicle rerouting, leads to an even larger reduction in the number of rejected passengers, as well as significant additional reductions in passenger travel and waiting times. The inclusion of the online dynamic insertion algorithm with demand forecasts increases the vehicle mileage only marginally, compared to using just empty vehicle rerouting.

## *5.2. Limitations and Future Studies*

One of the main limitations of the model is that a full distance and path matrix for the network are required as input. In larger networks (for example at a city level) it is perhaps prohibitive to store a complete distance and path matrix due to memory limitations. In such cases, a distance and path calculation needs to be performed at each insertion step, thus significantly increasing computation

times. Furthermore, the requirement of providing a distance and path matrix as input also hinders the possibility of including stochastic travel times. The modeling assumption that travel times can be considered deterministic, may be limiting in the case of a larger network, especially one where the (automated) vehicles share their infrastructure with other traffic.

The demand distribution is expected to be an important determinant of the efficiency of the rerouting as the demand forecast is only used for determining where trips might originate. The scalability of the algorithm should also be tested. Simulation times are determined by the size of the network, fleet size and the number of travel requests. Another limitation that will have an impact on the efficiency of the algorithm, especially on empty vehicle rerouting, is the fact that there is no vehicle capacity restriction at stops. Thus, an unlimited number of vehicles can stand idle at a stop. The impact of station capacity space limitations can be studied in the future by setting maximum capacity per station and incorporating it into the anticipatory vehicle assignment decision.

The operator costs were taken into account in the form of the vehicle mileage which is the main determinant of the variable costs of fully-automated vehicles. Since this research focused on the real-time operational aspect of vehicle rerouting, fleet size was assumed fixed. Further research may consider the ramifications of more efficient operations on tactical planning decisions such as determining the desired fleet size.

Future research may also investigate the potential benefits associated with vehicles with varying sizes, i.e. a heterogeneous fleet. Especially in the case where there are a few locations where a lot of demand originates, sending larger vehicles to those locations while using smaller vehicles for the other locations could be beneficial. This would of course require adaptations to the empty vehicle rerouting and insertion algorithms. Another adaptation to the current model, which uses only demand-responsive transport, is the combination of line-based transport and demand-responsive transport. It might be that using a line-based service for the stops with highest demand while using demand-responsive transport as a feeder service for the stops with low demand is more efficient than using exclusively demand-responsive transport.

- [1] N. Agatz, A. Erera, M. Savelsbergh, and X. Wang. Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research*, 223(2):295–303, 2012.
- [2] J. Anderson, N. Kalra, K. Stanley, P. Sorensen, C. Samaras, and O. Oluwatola. *Autonomous Vehicle Technology: A Guide for Policymakers*. RAND Corporation, Santa Monica, California, 2014.
- [3] R. W. Bent and P. van Hentenryck. Scenario-based planning for partially dynamic vehicle routing with stochastic customers. *Operations Research*, 52(6):977–987, 2004.
- [4] M. Bruni, F. Guerriero, and P. Beraldi. Designing robust routes for demand-responsive transport systems. *Transportation Research Part E: Logistics and Transportation Review*, 70:1 – 16, 2014. ISSN 1366-5545. doi: <http://dx.doi.org/10.1016/j.tre.2014.06.002>. URL <http://www.sciencedirect.com/science/article/pii/S1366554514000970>.
- [5] A. M. Campbell and M. Savelsbergh. Efficient insertion heuristics for vehicle routing and scheduling problems. *Transportation Science*, 38(3):369–378, 2004.
- [6] J.-F. Cordeau and G. Laporte. The dial-a-ride problem: models and algorithms. *Annals of Operations Research*, 153(1):29–46, 2007.
- [7] CROW. Kostenkengetallen Regionaal Openbaar Vervoer (Dutch). Technical report, CROW, 2015.

- [8] F. Ferrucci, S. Bock, and M. Gendreau. A pro-active real-time control approach for dynamic vehicle routing problems dealing with the delivery of urgent goods. *European Journal of Operational Research*, 225(1):130–141, 2013.
- [9] M. Gendreau, G. Laporte, and F. Semet. A dynamic model and parallel tabu search heuristic for real-time ambulance relocation. *Parallel Computing*, 27(12):1641–1653, 2001.
- [10] F. Guerriero, M. E. Bruni, and F. Greco. A hybrid greedy randomized adaptive search heuristic to solve the dial-a-ride problem. *Asia-Pacific Journal of Operational Research*, 30(01), 2013.
- [11] L. Hame. An adaptive insertion algorithm for the single-vehicle dial-a-ride problem with narrow time windows. *European Journal of Operations Research*, 209(1):11–22, 2011.
- [12] J. Holmgren. The efficiency of public transport operations: An evaluation using stochastic frontier analysis. *Research in Transportation Economics*, 39(1):50–57, 2013.
- [13] J.-J. Jaw, A. Odoni, H. Psaraftis, and N. Wilson. A heuristic algorithm for the multivehicle advance request dial-a-ride problem with time windows. *Transportation Research. Part B: Methodological*, 20B(3):243–257, 1986.
- [14] J. D. Lees-Miller. *Empty vehicle redistribution for personal rapid transit*. PhD thesis, Liverpool John Moores University, 2011.
- [15] J. D. Lees-Miller, J. Hammersley, and N. Davenport. Ride sharing in personal rapid transit capacity planning. *Automated People Movers*, pages 321–332, 2009.
- [16] O. B. G. Madsen, H. F. Ravn, and J. M. Rygaard. A heuristic algorithm for a dial-a-ride problem with time windows, multiple capacities, and multiple objectives. *Annals of Operations Research*, 60(1):193–208, 1995.
- [17] C. Mulley and J. D. Nelson. Flexible transport services: A new market opportunity for public transport. *Research in Transportation Economics*, 25(1):39–45, 2009.
- [18] J. D. Nelson, S. Wright, B. Masson, G. Ambrosino, and A. Naniopoulos. Recent developments in Flexible Transport Services. *Research in Transportation Economics*, 29(1):243–248, 2010.
- [19] M. Pavone, E. Frazzoli, and F. Bullo. Adaptive and Distributed Algorithms for Vehicle Routing in a Stochastic and Dynamic Environment. *IEEE Transactions on Automatic Control*, 56(6): 1259–1274, 2011.
- [20] V. Pillac, M. Gendreau, C. Guéret, and A. L. Medaglia. A review of dynamic vehicle routing problems. *European Journal of Operational Research*, 225(1):1–11, 2013.
- [21] L. Quadrioglio, M. M. Dessouky, and K. Palmer. An insertion heuristic for scheduling mobility allowance shuttle transit (MAST) services. *Journal of Scheduling*, 10(1):25–40, Feb 2007.
- [22] U. Ritzinger, J. Puchinger, and R. Hartl. A survey on dynamic and stochastic vehicle routing problems. *International Journal of Production Research*, pages 1–17, 2015.
- [23] D. O. Santos and E. C. Xavier. Taxi and Ride Sharing: A Dynamic Dial-a-Ride Problem with Money as an Incentive. *Expert Systems with Applications*, 2015.

- [24] M. Schilde, K. Doerner, and R. Hartl. Integrating stochastic time-dependent travel speed in solution methods for the dynamic dial-a-ride problem. *European Journal of Operational Research*, 238(1):18–30, 2014.
- [25] D. Tas, N. Dellaert, T. van Woensel, and T. de Kok. The time-dependent vehicle routing problem with soft time windows and stochastic travel times. *Transportation Research Part C: Emerging Technologies*, 48:66–83, 2014.
- [26] P. Toth and D. Vigo. Heuristic Algorithms for the Handicapped Persons Transportation Problem. *Transportation Science*, 31(1):60–71, 1997.