Delft University of Technology

Improvement of Network Performance by In-Vehicle Routing Using Floating Car Data

Klunder, Gerdien; Taale, Henk; Kester, L; Hoogendoorn, Serge

Research Article

Improvement of Network Performance by In-Vehicle Routing Using Floating Car Data

Gerdien A. Klunder,[1] **Henk Taale,**[2] **Leon Kester,**[3] **and Serge Hoogendoorn**[2]

[1]*TNO, Sustainable Urban Mobility and Safety, The Hague, Netherlands*
[2]*Department of Transport & Planning, Delft University of Technology, Delft, Netherlands*
[3]*TNO, The Hague, Netherlands*

Correspondence should be addressed to Gerdien A. Klunder; gerdien.klunder@tno.nl

This paper describes a study which gives insight into the size of improvement that is possible with individual in-car routing advice based on the actual traffic situation derived from floating car data (FCD). It also gives an idea about the required penetration rate of floating car data needed to achieve a certain degree of improvement. The study uses real loop detector data from the region of Amsterdam collected for over a year, a route generating algorithm for in-car routing advice, and emulated floating car data to generate the routing advice. The case with in-car routing advice has been compared to the base case, where drivers base their routing decisions on average knowledge of travel times in the network. The improvement in total delay using the in-vehicle system is dependent on penetration rate and accuracy of the floating car data and varies from 2.0% to 3.4% for 10% penetration rate. This leads to yearly savings of about 15 million euros if delay is monetarised using standard prices for value of time (VOT).

1. Introduction

By routing individual vehicles, it is clear that individual travel times and possibly also total network travel time can be improved. With recent technologies, a personal routing advice can be determined and presented to individual car drivers, by using an in-car device. This routing advice can be based on real-time traffic data, for example, floating car data generated by other drivers using the same service, which comprises experienced travel times. Recently, several pilots within the "Practical Trial Amsterdam" have been performed to test such a service [1, 2] in the Netherlands. However, relatively small implementation and compliance rates gave little insight into the potential effects for large-scale implementation, since route choice effects may positively influence travel times due to better utilization of available capacity, but only when a sufficient number of drivers will change their route.

In order to be able to determine a good routing advice based on the current traffic situation, adequate knowledge of this situation is necessary. Especially in the case of nonrecurrent, unexpected events, such as car accidents, routing advice can be very useful to shorten travel times. However, in order to detect such an unexpected event, sufficient and real-time traffic measurements need to be available. On Dutch motorways traffic measurements of loop detectors are available and of good quality, but much less traffic measurements are available for urban networks. For those networks floating car data can be used as an additional source. However, this raises certain questions. For example, what should be the amount and quality of FCD in order to be able to determine an adequate routing advice? And, what may be the improvement of such routing advice, both for the individual driver and for the network as a whole? This paper tries to answer these questions. First the research approach is presented; then the importance of determining the relation between quality of traffic data and the performance of a traffic management measure is explained. Next, the underlying data and the smart routing algorithm are described, after which the results are presented. Finally, the research questions are answered, conclusions are drawn, and recommendations for further research are given.

2. Importance of Determining the Relation between Quality of Traffic Data and the Performance of Traffic Management

Traffic management measures are designed with the (underlying) assumption that one has perfect traffic data available, which the traffic management system uses to operate. However, since in reality traffic data are never perfect, the traffic management measure will not perform optimally. But what effect can be achieved with these inaccurate or incomplete data and how much it differs from the "optimal" situation are usually not known. If these were known, it is, for example, possible to do a cost-benefit analysis on the goals the measure will achieve in relation to the costs of the data collection. Is it worth equipping more measuring points or more people with a measuring device? For ramp metering a start has been made to estimate the impact of inaccurate data and the benefits if cameras are used instead of loop detectors [3].

In our specific case, it is already interesting to find out how a measure as smart routing will perform in an urban network anyway, where most of the people already have some knowledge of the (regular) congestion. Another question could be how many data should be gathered (by floating devices) in order to obtain at least a positive effect. Finally, the effectiveness of route advice also depends for a large part on the normal route choice behaviour, the degree of succession of the advices, and access to other types of traffic information that the users already have. For this study these aspects are not taken into account, because this would require large-scale research on driving behaviour. However, the results can be interpreted for lower degrees of succession by looking at the results for lower penetration rates of the system.

3. Research Approach

To answer the research questions about the relation between the accuracy of data and the efficiency of the routing advice, a combination of modelling and using real (historical) data measurements has been chosen. The ground truth is derived from real traffic data measurements, while the quality variations of the FCD are modelled as perturbations of the real traffic data. Furthermore, since driven routes were not available (as is often the case), route choices have been modelled with commonly used mathematical models such as the logit model, as will be explained later. For a selected day, the historical speeds are used as input, while different situations for different FCD qualities and penetration rates of the smart routing system are modelled. To show the effect of different route choice options, several performance indicators are calculated on network level, in order to be able to quantify the effect of the differences in FCD quality.

Several modelling and data processing steps have been taken. The steps and the relation between them are shown in Figure 1. The first step was to know the ground truth of the traffic situation in the network. This means that the real speeds in the network during the investigation period should be available from measurements or estimations derived from measurements. For this, a large database with traffic measurements in the Netherlands was used. This traffic database was collected by the Dutch National Data Warehouse for Traffic Information (NDW, see [4]) and consists mainly of loop detector data. From this database the information for the region of Amsterdam was derived. A ground truth was created by using all available traffic data and using gap filling methods to complete missing data, as explained later.

The second step was to define zones in the network and to derive an origin-destination (OD) matrix with the number of trips from each origin to each destination for a given time period.

The third step was to determine reasonable alternative routes in the network between each OD pair, in order to be able to distribute the traffic over the network for the base situation and for the situation with smart routing. For each OD pair, a number of alternative routes were determined, based on aspects such as travel time, trip length, and road type.

These first three steps were preprocessing steps, independent of the smart routing system. The following steps were needed to test the smart routing measure for a selected day and for all variations of the FCD quality, time of day, and different penetration rates of the system.

Therefore the fourth step was the route selection, both for the normal users and for the users using smart routing. For the base situation of this study, without routing advice, a multinomial logit model was used to distribute the drivers over the route alternatives, based on the average network speeds. Link travel times were updated using calibrated BPR functions [5]. For the alternative situation using smart routing, the drivers were distributed either over only the shortest route or over three route alternatives with the shortest travel time (out of the predetermined set with route alternatives), using actual travel times and link speeds, as will be explained in more detail later. The network improvement was then determined for various penetration rates of the system and various times of the day and days of the year.

In order to determine the effect of the quality of the FCD, the amount of available FCD was varied by drawing a random number of links for which an actual speed measurement is assumed available. Links with a higher flow have a larger probability to be selected, since the probability that an FCD vehicle is found on these links is higher. This is accomplished by drawing a weighted sample with weights equal to the flow. For the links that are not covered by this data sample, the average (historical) link speed is used. Furthermore, in order to investigate the effect of inaccurate speed measurements, the quality of the floating car data was varied by adapting the link speeds with a random error. Both cases (variations in the amount and quality of FCD) were tested separately.

Finally, for varying implementation rates of the system, the improvement in travel times and delays (both individually and network-wide) was determined with regard to the base case and with regard to the optimal case based on perfect information.

4. Underlying Data

For the case study, we used different types of traffic data, such as loop detector data, travel times measured with cameras, and FCD, originating from the NDW database and
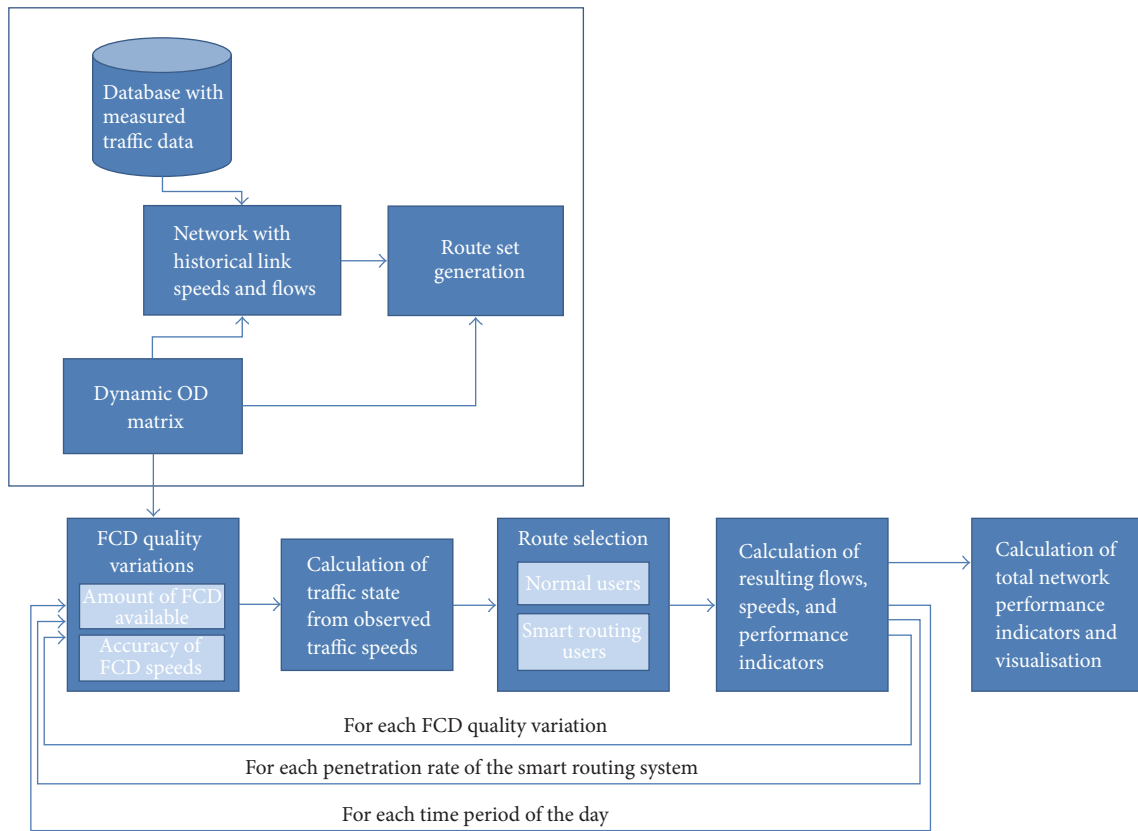
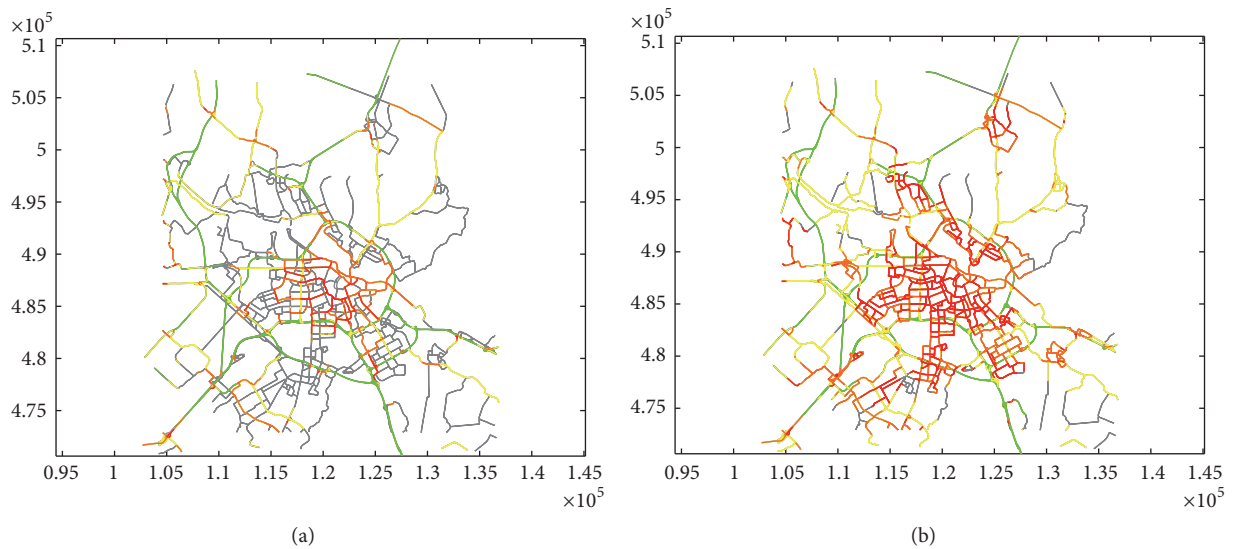FIGURE 1: Modelling approach for the smart routing case study.

(a)

(b)

FIGURE 2: Modelled network of the region of Amsterdam, before (a) and after (b) gap filling. The colours indicate the average speed: green > 80 km/h, yellow 50–80 km/h, orange 30–50 km/h, and red 0–30 km/h. For the grey links, no speed data could be derived.

the Practical Trial Amsterdam [6], fused into a database developed by TNO.

The network includes the city centre of Amsterdam and the surrounding region with a diameter of about 30 kilometres, as shown in Figure 2. It consists of 12,425 links. After gap filling and filtering, speeds and flows are available for most of the links for every minute of the day. For links for which no historical information was available, the gap filling method determines a link speed based on the available speed information of all links of the same road type within a range of 2 kilometres, with weights inversely proportional to the distance.
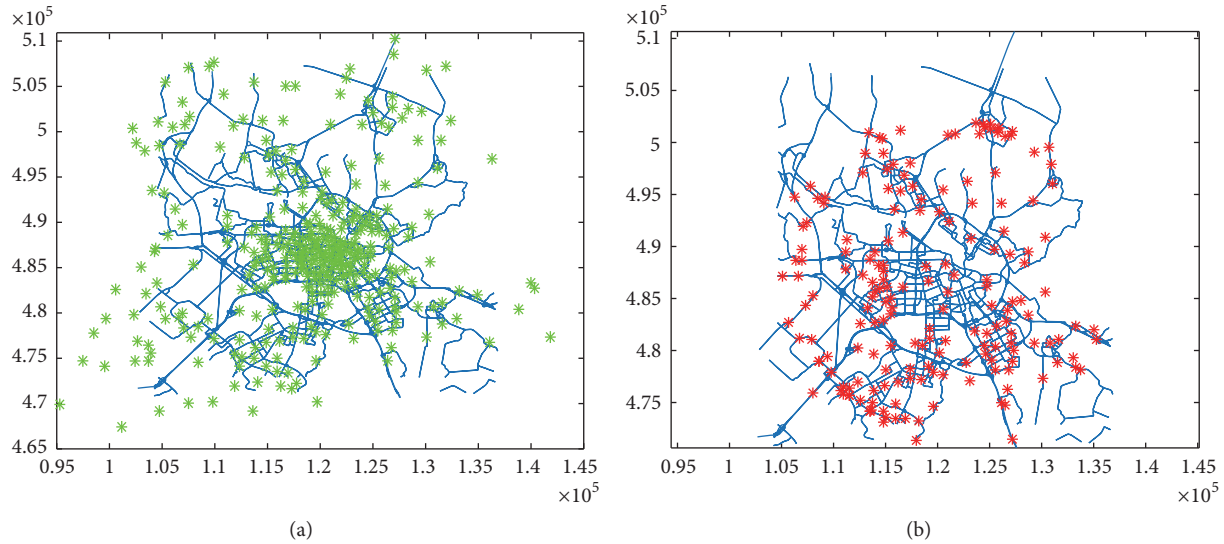
FIGURE 3: Zone aggregation in the network, (a) 350 zones before aggregation and (b) 183 zones after aggregation.

An OD matrix was obtained from a modelling study with the Regional Traffic Management Explorer [7]. This matrix came from a calibrated strategic model and was made dynamic by using departure time profiles, derived from inquiries among travellers. The matrix consisted of 350 zones with a high zonal density in the city centre, which is rather detailed for this small region. It turned out that this level of detail was not practical for this case study in terms of long calculation times and relevance for route choice. Therefore, the original matrix was aggregated into a smaller matrix of 183 zones. The (aggregated) zones also needed to be connected to the network (partly manually). The original zones (centres) and the aggregated zones are shown in Figure 3. A stronger aggregation was done within the city centre, because these zones are less relevant for route choice of cars, since most routes in the city centre are done by active modes (walking, cycling) and public transport. Most car trips in this network are through traffic on the major roads and to workplaces in the suburbs. Furthermore, much less information on real-time travel times from FCD (by car travellers) was available from the city centre.

The OD matrix was available for every quarter of an hour during the morning peak, between 05:30 and 11:00 hours, and for the evening peak between 14:30 and 20:00 hours. Because a continuous demand was needed for the daytime, for the quarters in between, the demand was estimated with a weighted average between each origin-destination pair. That means close to the end of the morning peak the last demand of the morning peak has a high weight and the first demand for the evening peak a low weight. The total demand over all origins and destinations for each period of the day is visualized in Figure 4.

5. The Smart Routing Algorithm

The smart routing algorithm in this case study consists of two parts, namely, an off-line route generation algorithm and an
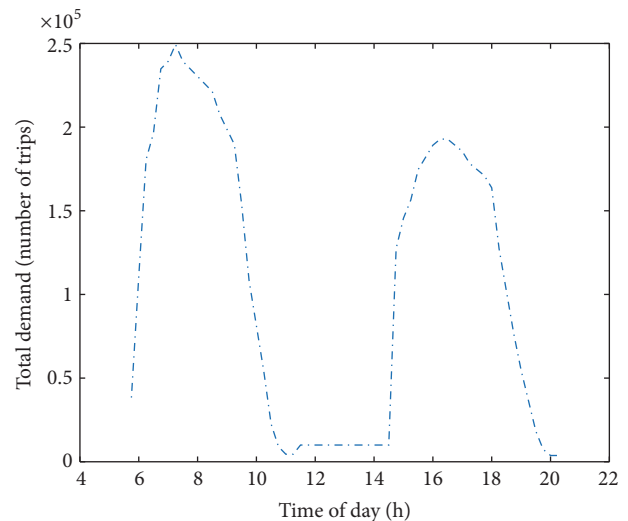

FIGURE 4: Total demand over the day.

on-line routing advice for individual road users. In practice, the off-line route generation is done as preprocessing step, while the routing advice is determined in real-time while a user is on the road. In this case study, both are done off-line in our data laboratory.

The off-line route generation algorithm generates route alternatives between each origin-destination pair, with the purpose of being able to distribute the traffic over these alternatives and in this way improve individual and/or total travel times in the network. The route alternatives should be good alternatives for the trips of the road user. They are calculated in advance, based on historic traffic data in the network. The route generation is done based on shortest path calculations with a generalized cost function. The attributes of the link cost that are taken into account are

(i) the travel time (the lower the better),

(ii) the total distance (the lower the better),

(iii) the safety/comfort, expressed as the distance on the underlying road network (not motorways) as percentage of the total route length (the lower the better),

(iv) the comfort, expressed as the average capacity per kilometre (the higher the better: sufficient capacity, more lanes).

Each of these attributes has a weight factor to determine the overall score of the route. These weighting factors can be varied per user or user type, as was done in the PPA [8]. Here we used fixed weighting factors $w_1 = 1$, $w_2 = 0.25$, $w_3 = 0.20$, and $w_4 = 0.0025$. In order to find route alternatives, first the fastest route, the shortest route, and the route with highest capacity are determined and added to the route set. In order to create more route alternatives, a Monte Carlo approach was used where the generalized costs per link are varied stochastically. The maximum number of routes that are generated per origin-destination pair is adjustable. In this case study, a maximum of ten routes was chosen. As will be explained later, all car drivers without the in-car routing device are distributed over all of these routes, while the users with the in-car routing device will be distributed over a maximum of three routes with the shortest real-time travel time.

The on-line routing advice will provide a real-time routing advice at departure time to individual road users. For each user, a selection is made from the ten route alternatives for his origin and destination. The pregenerated route alternatives are evaluated real-time with the measured travel times. Several strategies are possible. The easiest approach is to provide only the route with the shortest travel time to the user. Downside of this strategy is that if too many users receive and follow this advice, the capacity of this route might be exceeded such that this route is beginning to suffer from congestion. Another strategy is to provide a number of route alternatives to the user, from which he/she will make his own route choice, based on his personal preferences. Both strategies have been evaluated in this case study; for the second strategy the user was given a choice between 3 routes.

6. Accuracy and Availability of FCD Data

Accuracy of single FCD measurements is dependent on both the measurement accuracy of the device (e.g., GPS or WIFI and quality of the GPS receiver), the environment (e.g., high density of high buildings will deteriorate the accuracy of GPS), and the software that is used for, for example, map-matching, speed calculation, and corrections. The average position error ranges from 2 meters on an open square to 15 meters in wide streets with buildings on both sides [9]. For speed calculation a GPS usually takes a running average of data points with some smoothing function. This means that while accelerating or decelerating the GPS will have a larger error than at constant speed. Besides location measurements, signal Doppler shift is also used to make it more accurate. Speed measurements can furthermore be extracted in other ways, for example, directly from the car; however, speed
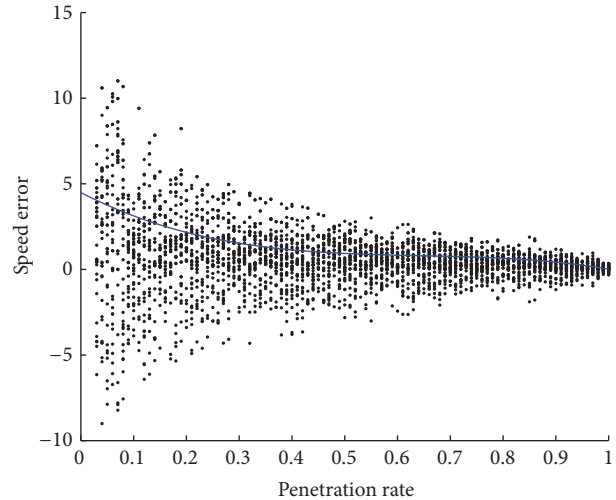

FIGURE 5: Estimation of percentual speed errors for the average speed at a road section with speed limit 80 km/h, moderate traffic flow, and a length of 400 m, based on FCD samples with different penetration rates. The solid curve represents the average absolute percentage error.

extracted from GPS is usually more accurate than measured by the car, since it is not affected by inaccuracies such as the vehicle's wheel size or drive ratios. It is dependent however on GPS satellite signal quality but these errors can be minimized with the use of moving average calculations. However, GPS is more dependent on the environment than speedometers in the car, think of tunnels, high buildings, and so forth. For GPS speed accuracy based on GPS-Doppler 10-second average speed accuracy better than 5 cm/s with the confidence level better than 99.9% has been observed [10].

One could conclude that single GPS speed measurements are fairly accurate, though, for average speed calculations of a road section the accuracy of the average speed depends more on the availability of FCD than on the accuracy of single FCD speed measurements. Let us do a simple calculation exercise. For example, assume a road section of 400 meters with a speed limit of 80 km/h and a flow of 1000 veh/h, which corresponds to a density of 18 vehicles on this road section. Assuming furthermore that the real speed of these vehicles is normally distributed with mean 82 km/h and a standard deviation of 5 km/h and that the measured FCD speed has a normally distributed error with a mean of 1 km/h, the FCD speed error (difference between real average speed on the road section and the average measured FCD speed) for different FCD penetration rates is shown in Figure 5. A similar calculation is done for the network of the region of Amsterdam, separately for each link in the network, considering the speed limit, link length, and estimation of the flow, for a penetration rate of 1%, 10%, 50%, and 90%. The result is given in Table 1. For 1% the density is in most cases too low to be able to estimate the error; therefore no estimation could be made for this penetration rate.

In addition to the speed accuracy, low FCD penetration rates have an additional problem; namely, during the

measurement period at (short) road sections not any FCD vehicle might be present. The probability for this can also be calculated statistically; assuming constant speeds one can use the Poisson distribution for this. This probability is furthermore dependent on the duration of the time interval, the flow (q), the link length, and the average speed. Vehicles present on the link at the beginning of the interval are also included by extending the time interval with the average travel time. The results for a three-lane motorway section (100 m) and an urban road section are shown in Figure 6 (notice the difference in axes scale). Notice that, on the motorway with moderate and high flow, already above 2% penetration rate the probability that there is no FCD available is negligible. On the urban road, this is around 20%.

However, often one wants to have more than one observation on a link in order to calculate an average speed which is accurate enough (to average out speed dynamics on the road section and measurement inaccuracies or to remove odd cases such as vehicles resting at a parallel parking space). The same calculation can be done for the probability of at least 3 FCD vehicles to be present on a road section during a given time interval. The results are shown in Figure 7. This led to quite different results; on the motorway the probability that there are less than 3 FCD available is negligible above 4% penetration rate and on the urban road above 40% penetration rate.

Taking this exercise further to estimate the relationship between the penetration rate of FCD vehicles and the availability of reliable average link speeds in a traffic network, we used this calculation method together with the network data of our case study around Amsterdam with measured and estimated speeds and flows for all links (12,425 links, average link length 186 m), on the January 28 at 8:00, assuming a uniformly distributed FCD penetration rate throughout the network. For each link, the probability that sufficient FCD is available is estimated and summed up for the whole network. From this, the expected number and percentage of available links are calculated. The result is shown in Figure 8. Relating the link availability percentages to FCD penetration rates, a link availability of 10%, 50%, and 90% corresponds to, respectively, 1.7%, 5.8%, and 15.6% FCD penetration rate throughout the network. From this one can conclude that already quite low FCD percentages lead to relatively high availability of links where a reasonably accurate average speed can be calculated. The other way around, looking at the FCD penetration rates as used later in this paper, we get the percentages as given in Table 1.

7. Modelling Smart Routing

In this paragraph, we explain in more detail how the route choice was modelled for both "normal" users (those who do not have the smart routing system to their disposal) and smart routing users, as well as which days were modelled and how the quality variations in the traffic data are modelled.

7.1. Modelling Normal Users. We assume that the "normal users" do not have real-time information about the actual

TABLE 1: Relation between penetration rate of FCD, speed error, and link availability.

Penetration rate of FCD	Speed error (%)	Link availability (%)
1%	NaN	5%
10%	5.6%	76%
50%	4.6%	99%
90%	1.1%	99.5%

speeds on the road network but that they have a notion of what the speeds usually are on their routes, derived while driving in the network during the recent past. In the model, this is estimated by calculating the mean speeds on each link in the network over the last two months. The route choice of these users is based on the average travel times on their routes from these average speeds (without the speed of the current day). The route choice is modelled with a multinomial logit model. The multinomial logit model (MNL) is the most widely used choice model, due to its simple mathematical structure and ease of estimation [11]. In this model, the probability for using route r is calculated as follows:

$$P(r) = \frac{e^{-\theta T_r}}{\sum_r e^{-\theta T_r}} \tag{1}$$

in which T_r are the generalized costs (in this case the estimated travel time) of the route and θ is a scaling parameter which represents the knowledge level of the user. In this case study, we used the following settings: theta = 1, T_r travel time in minutes, which has been proven in earlier studies to be realistic settings [12]. More sophisticated route choice models exist which, for example, correct for overlapping of routes (see [11, 13, 14]). This case study focuses on the effect of inaccurate input data rather than on route choice modelling. A route choice model that takes overlapping of routes into account could also represent the effects of inaccurate information on route choice better; however, in the current study this is not (yet) done, since this would lead to much longer calculation times, while the calculation time is already a limiting factor in this study due to the large number of variants in information levels. Also, the route generation model already filters out routes that greatly overlap. Therefore it was accepted that the multinomial route choice model is sufficient for the purpose of this study.

The route flows for every OD pair and every time period are then calculated by multiplying the probabilities with the number of trips per OD pair.

7.2. Modelling Smart Routing Users. Between each OD pair, we derived at most ten alternative routes, as mentioned in Section 5. A certain fraction of all drivers uses the smart routing application, which we call the penetration rate p of the system. The follow-up rate of the system is already contained in this fraction, but can be modelled separately when desired. Users with the smart routing application will get a routing advice representing one or more routes to their destination. Here we investigated first the scenario that only
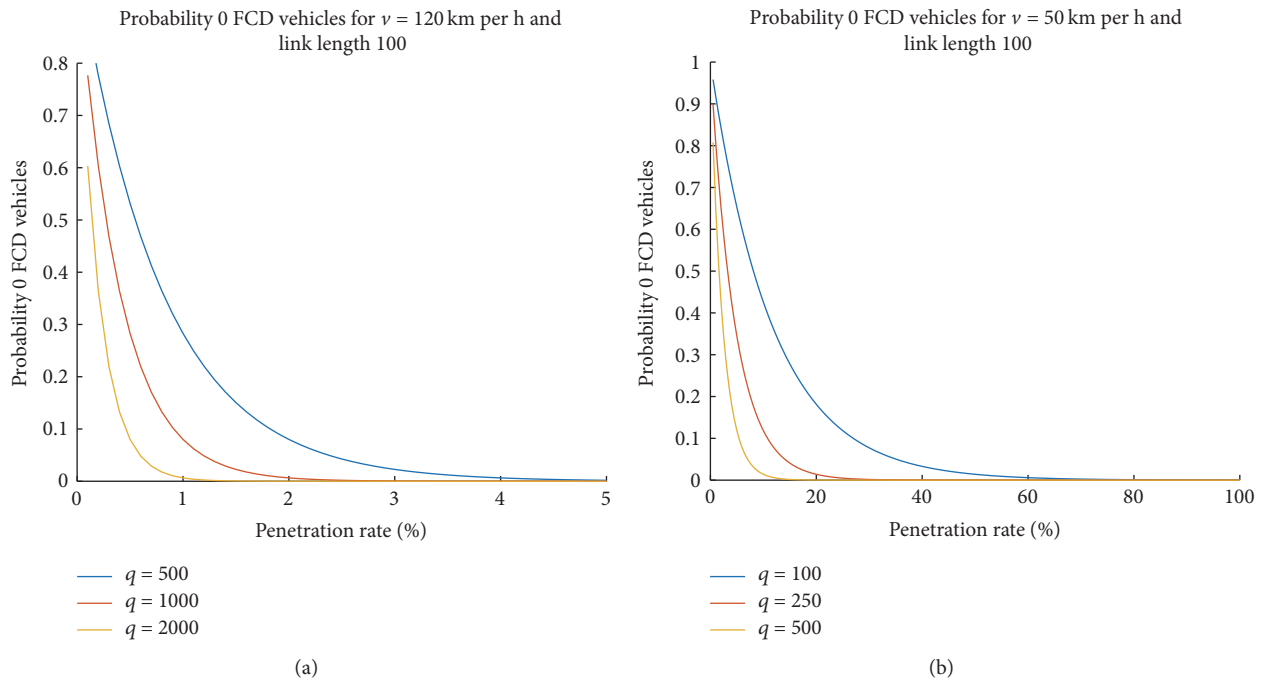
FIGURE 6: Estimation of the probability that not any floating car is observed during a 5-minute time interval on a road section of 100 m length and different lane flows (q), on a 3-lane motorway with speed limit 120 km/h (a), and on a one-lane urban road with speed limit 50 km/h (b).
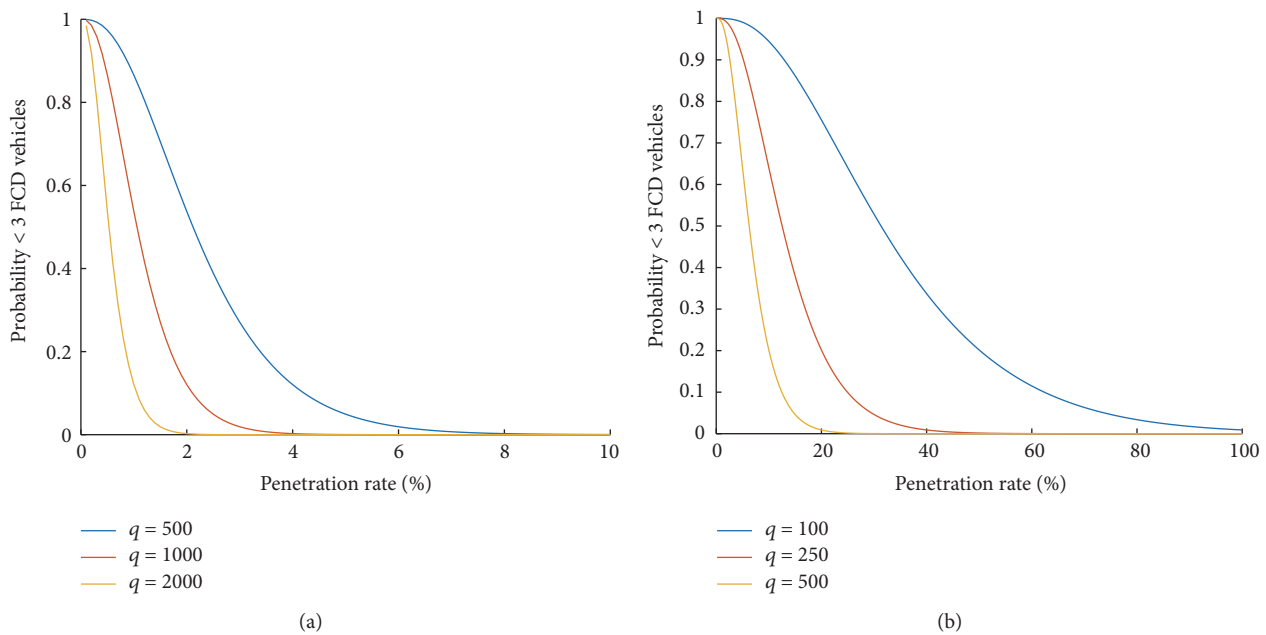

FIGURE 7: Estimation of the probability that less than 3 FCD vehicles are observed during a 5-minute time interval on a road section of 100 m length and different lane flows (q), on a 3-lane motorway with speed limit 120 km/h (a), and on a one-lane urban road with speed limit 50 km/h (b).

the route with the shortest actual travel time is advised to the users, which they will accept. Secondly, we investigate the scenario that the three best/fastest routes based on the actual travel time are shown to the user. The users route choice is assumed to be distributed over this top 3 conform the multinomial logit model, similar to the users without smart

routing advice, but in this case based on the actual speeds instead of average speeds. The routing advice does not take into account the capacity and actual flows on the routes. This is to conform most current routing advice systems work, since it is a difficult task to estimate actual flows in the network and to determine a routing advice based on, for example, a user
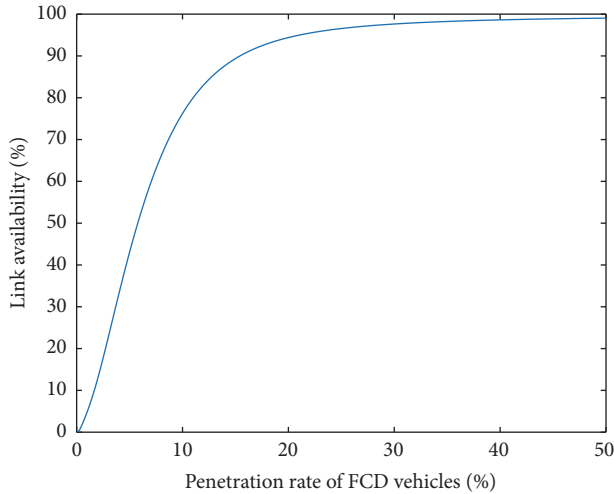
FIGURE 8: Estimation of the link availability for different penetration rates of FCD vehicles on the study network of the region of Amsterdam.

equilibrium in order to prevent that too many drivers choose the same route leading to congestion. But we do take into account the fact that the resulting flows might lead to longer travel times when certain route parts approach or exceed their capacity, using the so-called BPR function [15]:

$$
\text{tt}_{\text{cong}} = \text{tt}_{ff} \cdot \left(1 + \alpha \left(\frac{q}{C} \right)^{\beta} \right). \tag{2}
$$

The BPR function is calibrated on the speed data in the network on link level (for each link separately) with a polynomial fit, for the links for which measured speeds were available. Capacity of the links was estimated based on the speed limit and the number of lanes. For the other links, default values for alpha and beta were used: alpha = 0.15 and beta = 4.

7.3. Selection of Days. The smart routing advice is expected to have a larger impact when the traffic situation is different from the daily traffic pattern, that is, when there is more congestion than normal, because then it has more added value to know the actual traffic situation. Therefore, we want to test the system for one or more days where the traffic speeds have a large deviation from the speeds on an average day, and for comparison also for a day which has speeds close to the ones on an average day. In order to find such days, the average network speed was calculated for each period of the day and all days in the first three months of 2015 for which sufficient data were available. Next, the average network speed was calculated over all weekdays (Monday to Friday). Days with a lot of outliers were excluded from the calculation of the average network speed. For each day, the network speed was compared with the average speed and a choice was made for an average day and for two (different) abnormal days. One of the chosen abnormal days is Tuesday, February 3, 2015. On that day there was a lot more congestion than usual. This was caused by slipperiness due to the snowy winter weather during the whole day. The network speed of the "abnormal"

day together with the network speed of the average weekday is shown in Figure 9(b), where it is shown that the actual network speed is much lower than the average network speed. The other abnormal day was the 5 February 2015. On that day there was also a lot more congestion than usual, caused by several incidents, especially during the morning peak, as shown in Figure 9(c). The selected average day is Wednesday, January 28, as is shown in Figure 9(a). A remark needs to be made that in exceptional days like those investigated, demand and capacities are different than normal days, and also uninformed users' behaviour may be different, since they will face different delays. However, we assume that most users will still take the route that they are used to take and therefore the estimated route choices are considered representative also for these cases.

7.4. Variation in Quality of the FCD. Floating car data is never 100% correct. In this study we want to know what the effect is of different levels of quality of the floating car data. There are different types of errors or inaccuracies in floating car data, but in this study we focus on two types of inaccuracies: lack of data and inaccurate data.

For the first inaccuracy type (lack of data) the number of links for which the actual traffic speed is known is varied. To be able to generate a complete advice, for the links for which no information is available, the free flow speed of the link is used. The percentage of links for which no information is available is varied from 0% to 100% with steps from 10%. The links for which no actual traffic speeds are available are drawn randomly with a higher weight for links with a higher flow. The flows are used as weights in a drawing without replacement. This is done in this way because in practice there is also a higher probability that floating car data are available on links where more vehicles have passed. The smart routing application will base its routing advice on the adapted, incomplete information of the speeds in the network. Since part of the congestion is not observed, these travel times will generally be shorter than the real travel times. However, it may still be better than the average travel times that are used for the route choices of the normal road users without the smart routing application.

For the second type of inaccurate data, we assume that the inaccuracy of the speeds found in the floating car data is normally distributed with mean value of the real average speed and a standard deviation that is varied from 0 to 0.9 with steps of 0.1, as shown in Figure 10. For each link, a separate drawing is done and applied to the speed of that link. The smart routing application will again base its routing advice on the adapted (inaccurate) information of the speeds in the network. The resulting travel times may be either larger or shorter than the real travel times, but again may still be better than the average travel times that are used for the route choices of the normal road users without the smart routing application.

8. Results

8.1. Results Using Perfect Information. Suppose we have access to perfect traffic information and we offer the smart routing
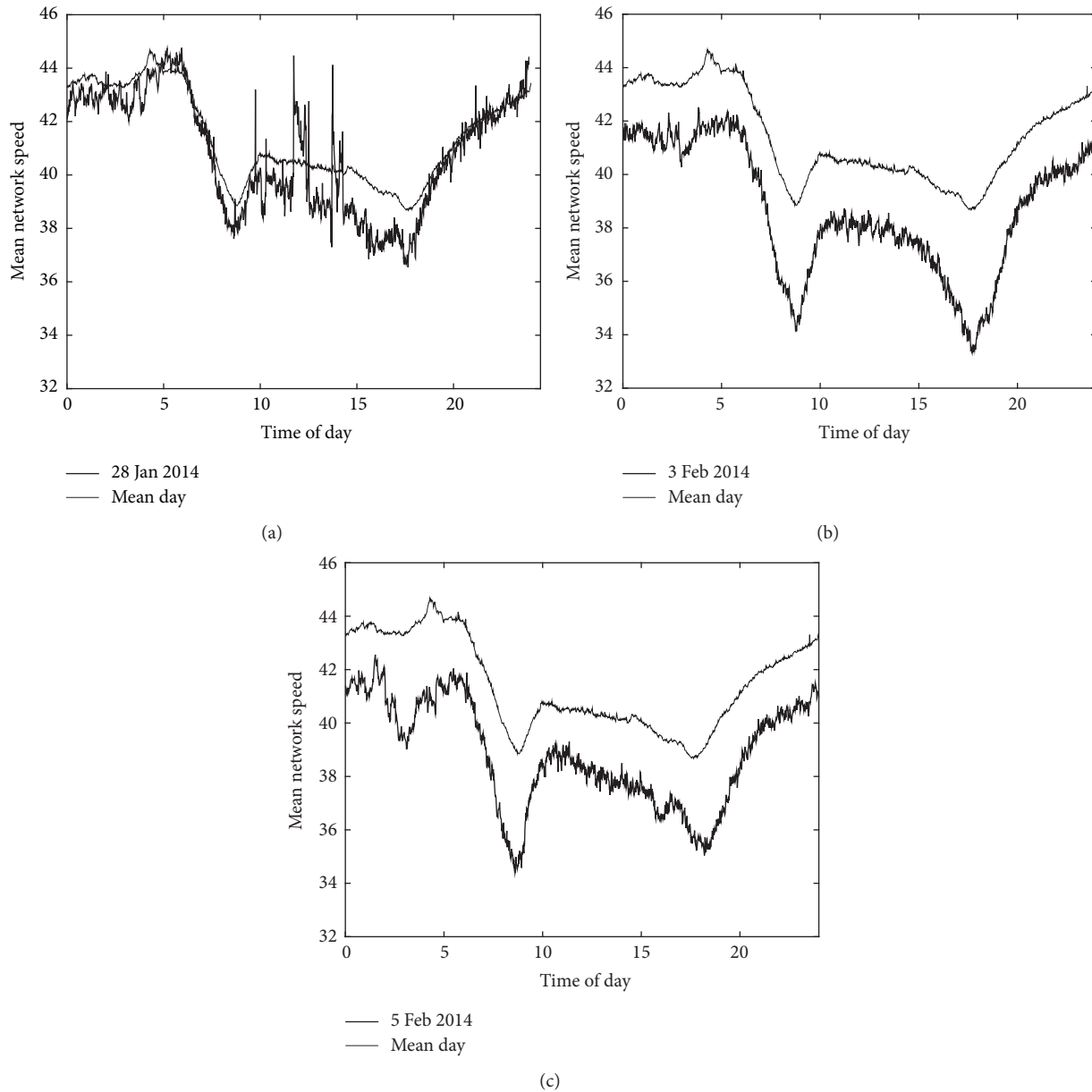
(a)

(b)

(c)

FIGURE 9: Selected day with an average speed pattern (a), a day with an abnormal speed pattern February 3, 2015 (b), and a day with an abnormal speed pattern on February 5, 2015 (c).

advice as explained above. Then we can calculate what the potential effect is of the smart routing application. Since the system knows the actual travel times, the total travel time in the whole network is expected to decrease, as well as the delays in the network. On individual level, smart routing users will benefit as well. Only when too many road users choose the same route, based on the advice, may (individual) travel times increase.

We have calculated the effects for different penetration rates of the system for the three different days. The total delay was calculated as the difference between the total of all free flow travel times in the network and the total of all actual travel times in the network. The difference in total delay without smart routing and with smart routing is compared

for several penetration rates of the system (1%, 10%, 50%, and 90%). The result is shown in Figure 11. This figure clearly shows that the improvement of the system is largest for high penetration rates of the system and during the peak hours, as can be expected.

In Figure 12, the total delay improvement over the whole day is shown for the three different days and the two route advice strategies (showing only the best route or the top 3 best routes). In Figure 12(a) it can be clearly seen that the day with the most deviating congestion (February 5) has the largest benefit in saving delay hours, as expected. It is striking that the results show a clear linear relationship with the penetration rate. It is difficult to understand this relationship; it seems that, despite the complexity of the network and the
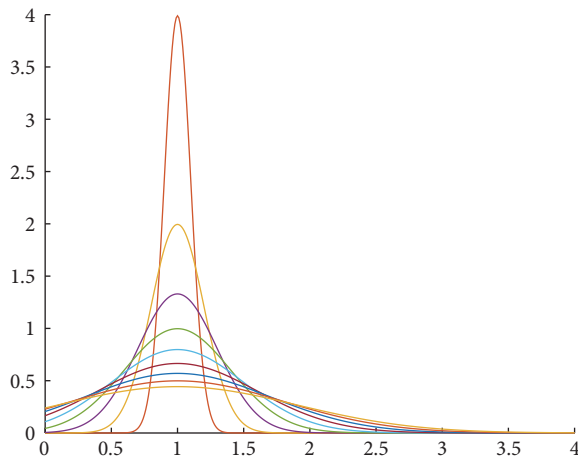
FIGURE 10: Error factors for the inaccuracy of the measured speeds.

nonlinearities in the overall system, the impact of more FCD in the system can be approached with this linear relationship, though it could also be the case that the way of modelling and the assumptions used oversimplify the outcomes. We however still believe that this will not change the general conclusions, order of magnitude of the effects, and other general insights gained from this study.

For a penetration rate of 10% on the average day with the best routing advice, the total improvement over the whole day is 4480 hours, or 0.8% of the total delay in the network as shown in Figure 12(b). Giving a top 3-route advice gives a little bit less good results of 3150 saved hours or 0.8%. Apparently the possible prevention of congestion on the best route does not compensate the longer travel times of the second and third best routes. For 90% penetration rate, 4.7% to 7.0% of the total delay in the network would be saved on the average day. It is also striking that the results of the average day and the day with the most congestion do not differ much percentagewise for the best routing advice only (8.6%).

Using a value of time of EUR 12.28 (assuming an average of 10% freight traffic and the key figures from [16]), this means a saving of 55,000 euros for the average day. This can be scaled up to a yearly level by calculating with the yearly number of weekdays (261 in 2015). This gives an estimate of 14.5 million euros saved on a yearly basis for 10% penetration rate of the system, when perfect traffic data and traffic information would be available. For 90% penetration rate 130 million euros could be saved on a yearly basis, calculating only with average days. However, since part of the days has more congestion than average, the potential is larger.

8.2. Results with Variation in Quality of the FCD. In Figure 13, the relative improvement in delay time compared with no routing advice is shown for the average day (January 28, 2015). (a) and (c) are the results where the system only gives the best route option to all users. In (b) and (d), the results when the 3 best routes are given are shown. In (a) and (b), the accuracy of the speed information is varied, while the graphs in (c) and (d) show the results for the variation of available link speeds.

As can be seen from these results, it seems that inaccurate speed information has a higher impact on the results than incomplete information, though these are different quantities and cannot be compared exactly. Furthermore, giving the user only the best route option leads to better results than giving the user a choice between the three best route options. The same as in the case with perfect information, it appears that the possible prevention of congestion on the best route does not compensate the longer travel times of the second and third best routes. However, if the errors on the link speeds are higher than 40%, the total impact becomes negative.

For the best route option only, the highest impact is around 7% with 90% penetration rate of the system. For the three route options, the impact stays below 5%. With 10% penetration rate, the impact is below 1% and around 0.5% with three route options. For 1% penetration rate, the effects on network level are negligible.

The completeness of the information has less (negative) influence, the impact stays within 4%–7% for 90% penetration rate of the system and 1 route option, even when no actual link speeds are available. Apparently the route advices based on estimated link speeds (free flow speeds) are still better than the route choice based on the average knowledge of drivers of speeds in the network.

Figure 14 gives comparable results for February 3, 2015, where more congestion was present in the network. Surprisingly, the impacts are almost comparable as for the average day. The cause of this can be that this day the congestion was due to snow and ice during the whole day and on all roads. Therefore not much can be gained with route advices for alternative routes. Only the link availability is more important on this day, for low link penetration rates.

Finally, the results for February 5, the most "abnormal" day, is shown in Figure 15. It is clear that the effects for this day are the highest; for the speed accuracy this is 8.6% for 90% penetration rate and 1% for 10% penetration rate and when you calculate the effect during the morning peak hour the effect is about 14% for 90% penetration rate and about 1.5% for 10% penetration rate.

Assuming that the drivers that use the smart routing app also provide the FCD data, the penetration rate of the users is the same as the penetration rate of the FCD, such that the percentages of Table 1 apply. These can be linked to the results of this paragraph. For the link speed error, a translation is needed from average absolute error to the standard deviation of the error distribution. For the percentages of Table 1 (5.6%, 4.6%, and 1.1%), this is, respectively, 0.07, 0.06, and 0.01. Linking this to the results of this paragraph, the FCD penetration rate hardly makes any difference for the results based on speed error or for limited link availability, as shown in Table 2. One should however take into account the fact that the link speed error is caused by more than a small random error for GPS inaccuracy, the data may also contain outliers in the speed by other causes, or in a densely built up area the error will be larger.

Link availability and speed inaccuracy are linked in the sense that a certain minimum availability of FCD is needed on a link in order to be able to actually calculate an average speed. The higher the availability, the higher the speed
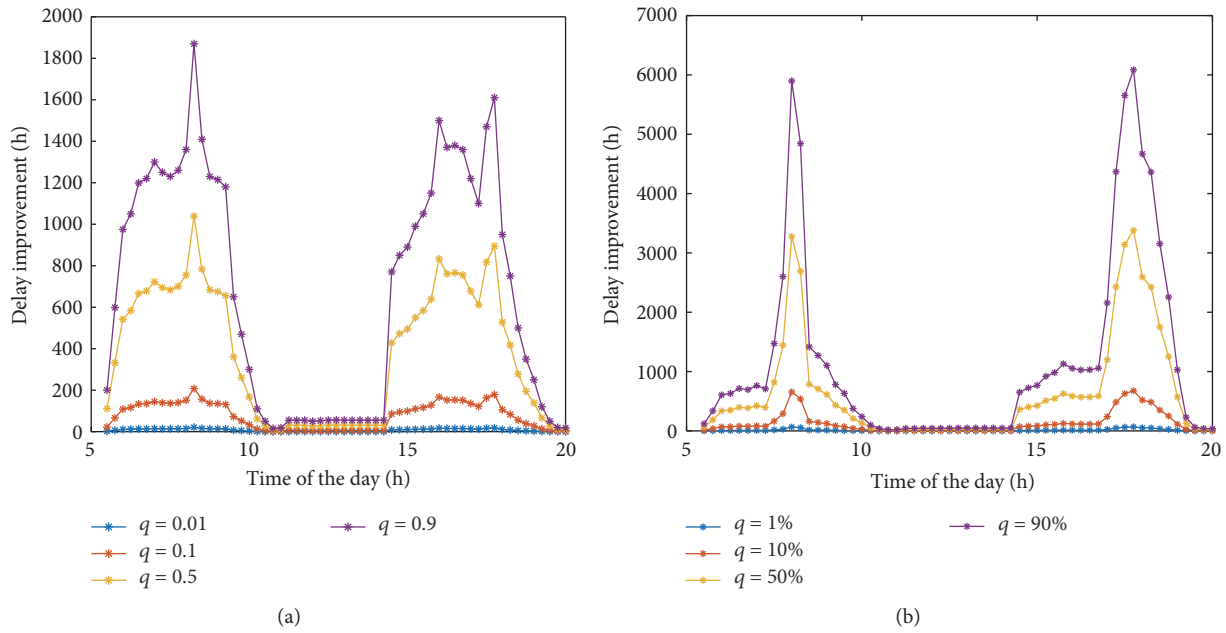
FIGURE 11: Effect of smart routing using perfect information for different penetration rates on the average day (a) and for February 5, 2015.
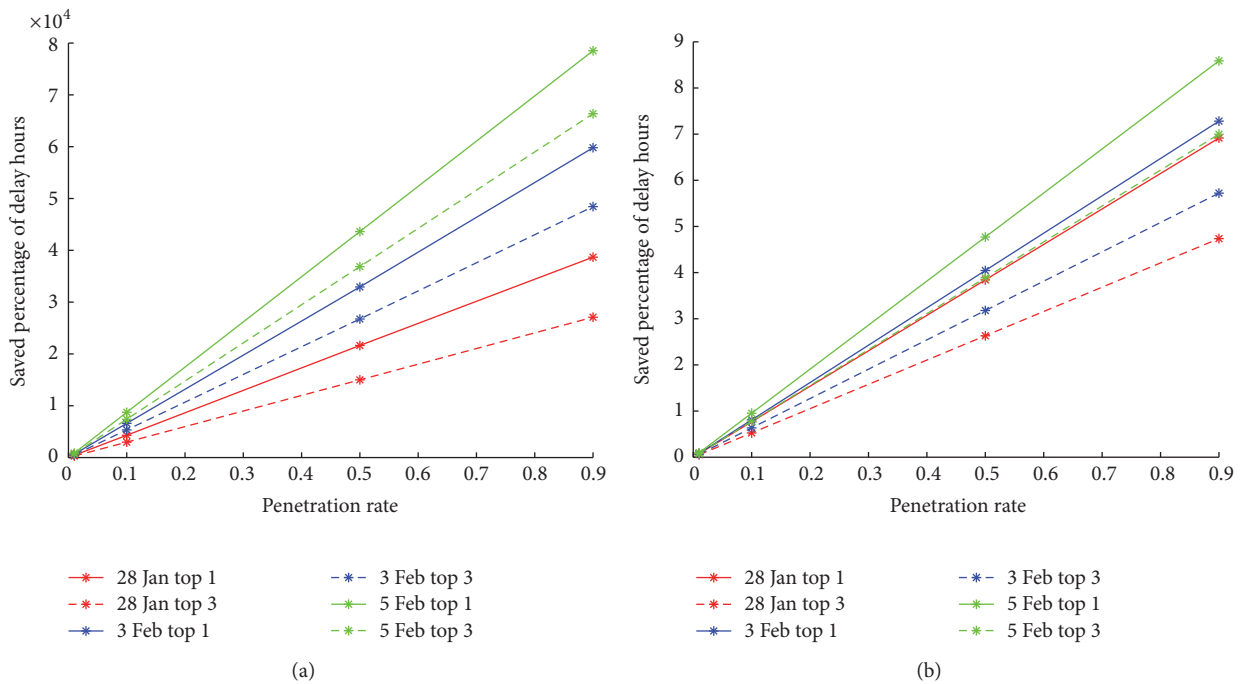

FIGURE 12: Saved delay hours as total over the whole day (a) and as a percentage with regard to the situation without routing advice (b).

TABLE 2: Relation between penetration rate of FCD, speed error, and link availability.

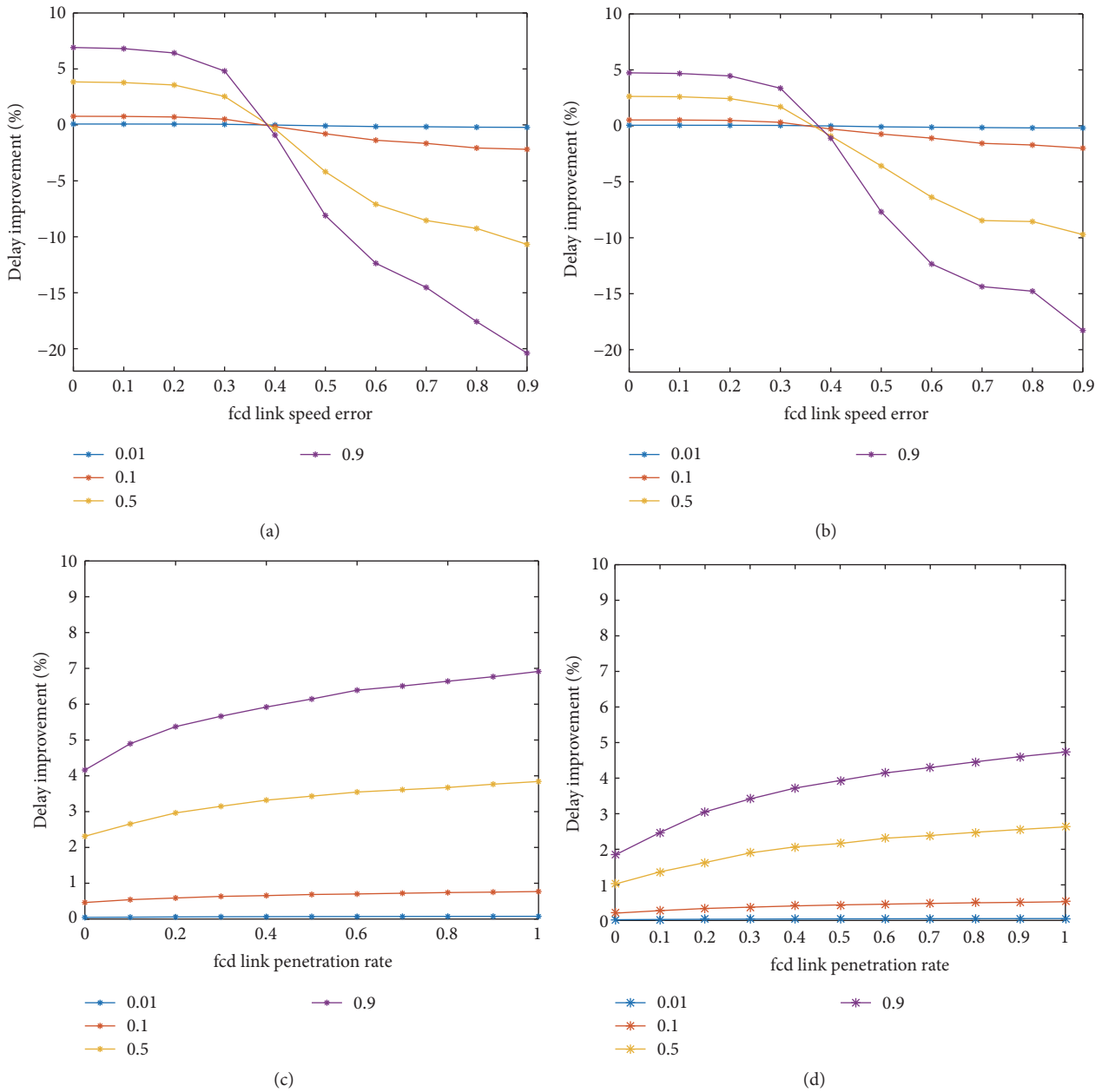| Penetration rate of FCD	Delay improvement for speed inaccuracies			Delay improvement for limited link availability		
	28 January	3 February	5 February	28 January	3 February	5 February
1%	Unknown	Unknown	Unknown	<0.03%	<0.02%	<0.02%
10%	0.8%	0.8%	1.0%	0.7%	0.7%	0.9%
50%	3.8%	4.0%	4.8%	3.8%	4.0%	4.8%
90%	6.9%	7.3%	8.6%	6.9%	7.3%	8.6%

FIGURE 13: Delay improvement for 28 January 2016 (average day) as percentage with regard to the normal delay: variation in speed accuracy with best route advice (a), variation in speed accuracy with top 3-route advice (b), variation in link speed availability with best route advice (c), and variation in link speed availability with top 3-route advice (d).

accuracy. For 50% and 90% FCD, the results are almost equal, because they are both very close to the maximum effect. An FCD percentage somewhere between 1% and 10% would show larger differences. Actually one should take into account both quality aspects in order to estimate the total effect; the combined effect is smaller than when taking into account only one of the inaccuracies.

9. Relation with Data Quality and Cost Benefit

Based on the results this case study allows for a cost-benefit analysis which relates the amount and quality of the input

(floating car) data to the improvement of the individual and network performance. This may be used to support investment decisions on such a system to answer questions such as what can be the benefit of real-time in-car routing, which implementation rate should be achieved in order to reach a certain improvement, and which quality of input data is needed (either from the participating vehicles or from another source)?

The benefits of the system are calculated for different penetration rates and qualities using a value of time in the same manner as above for the case with perfect information. This gives pictures of similar shape of Figures 13–15, up to
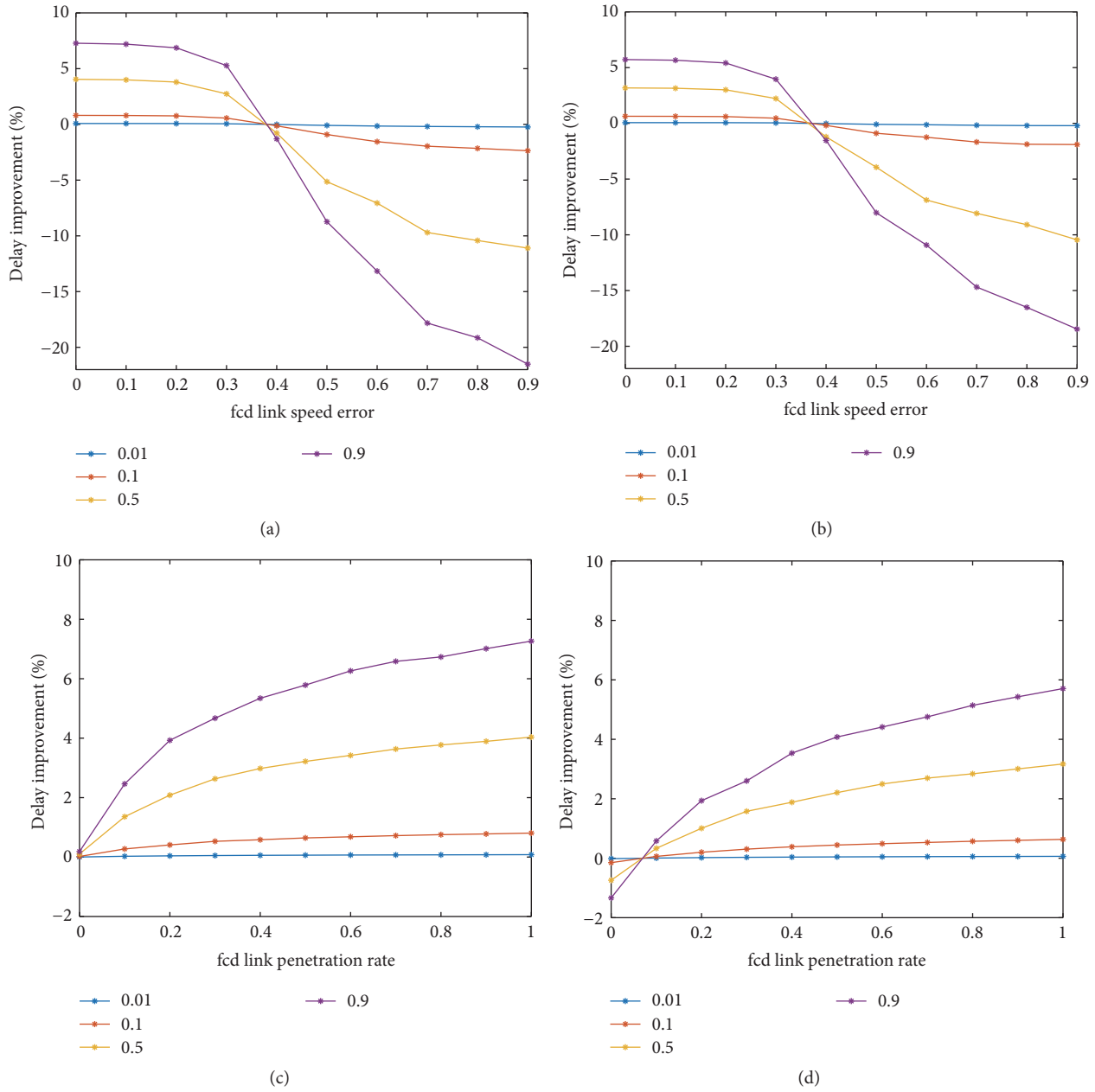
FIGURE 14: Delay improvement for 3 February 2016 (day with a lot of congestion) as percentage with regard to the normal delay: variation in link penetration with best route advice (a), variation in link penetration with top 3-route advice (b), variation in speed accuracy with best route advice (c), and variation in speed accuracy with top 3-route advice (d).

one million euros a day for 90% penetration rate on February 5.

The benefits can now be weighed against the costs of such a system and data gathering. These costs consist mainly of the software development, maintenance, and data communication costs. The data communication costs depend on the penetration rate of the system and on the sample rate: the higher the penetration rate and the higher the sample rate are, the more the communication is needed. However, in the case of a mobile phone service, nowadays most people have a fixed cost subscription with a large amount of data

communication included. Though we can assume that some quality aspects of the data are related to certain costs.

The total number of trips during one day in or crossing the region around Amsterdam in this case study is 6.4 million, based on the OD matrix (night not included, see Figure 4), or 2 million during the morning peak. Since most people make more than one trip a day, namely, 2.7 on average [16], the total number of different people making one or more trips in this area can be estimated at 2.4 million on a daily basis. Hence a penetration rate of, for example, 10% relates to approximately 240 thousand people (selected from the people
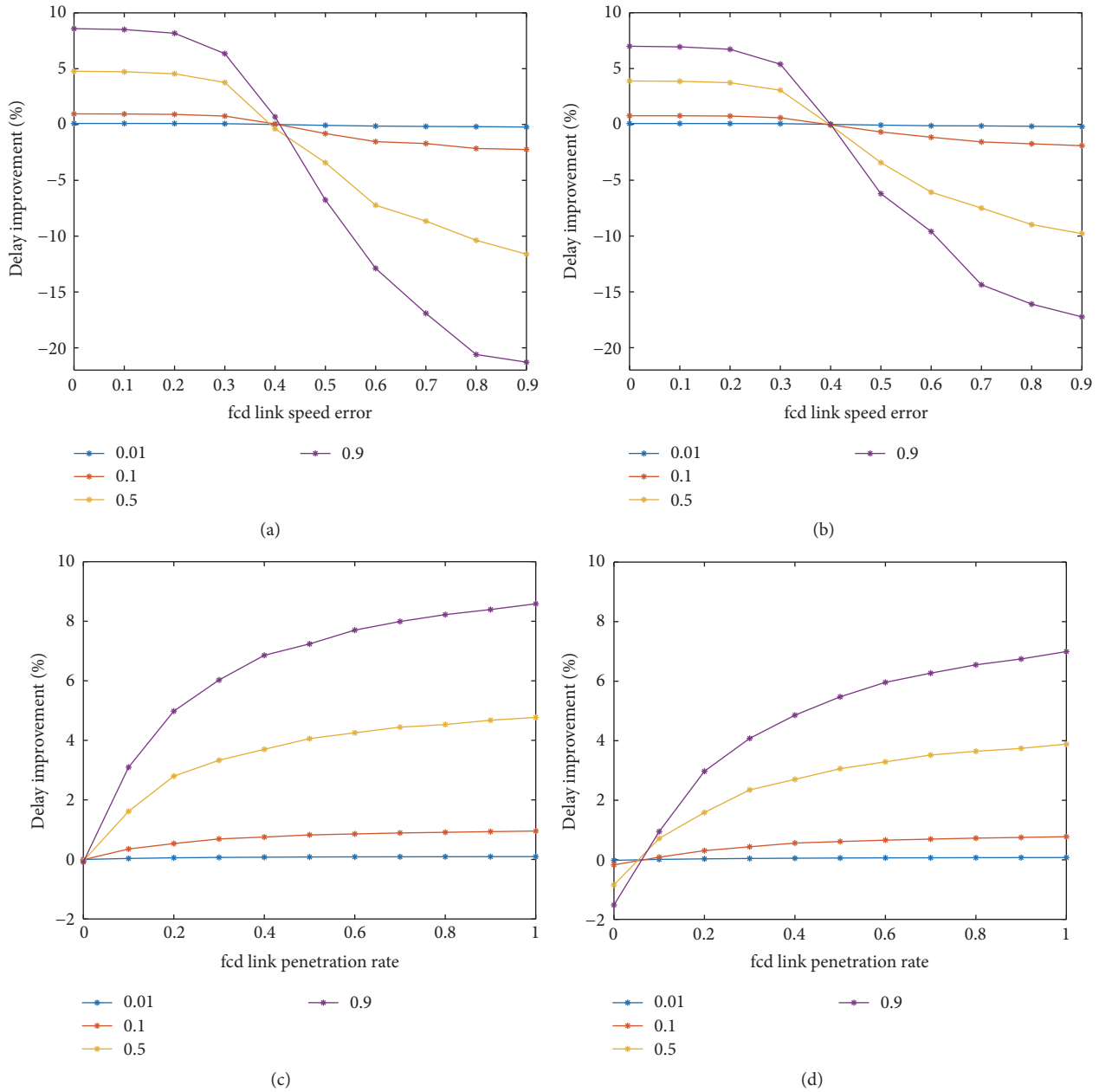
(a)

(b)

(c)

(d)

FIGURE 15: Delay improvement for 5 February 2016 (day with a lot of congestion) as percentage with regard to the normal delay: variation in link penetration with best route advice (a), variation in link penetration with top 3-route advice (b), variation in speed accuracy with best route advice (c), and variation in speed accuracy with top 3-route advice (d).

that are known to travel in or through this area). For the morning peak this is comparable, assuming that one usually makes only one trip during the morning peak, one gets 200 thousand people for 10% penetration rate. Assuming that they all need a mobile phone subscription with data for an average cost of 20 euros per month, this gives a total cost of around 200 thousand euros per working day. This does not weigh against the benefit in value of time per day that was calculated earlier (55000 for a normal working day); however, since most people already have a mobile phone subscription with data connection, the extra costs for this service are zero.

Additional costs can be considered for the data collection, analysis, and processing, which consist of a fixed part and a part dependent on the amount of data. Twice as much data usually needs more processing and analysis time, but not twice as much time, since certain analysis steps are as much work for a small dataset as for a large dataset, and independent calculation steps can be parallelized. The same scripts can often be used to handle a larger dataset. So assume, for example, that every month a fixed amount and a variable amount of man hours are used to handle the collected data correctly, say 10 hours (fixed) plus one hour for

the data of every 10 thousand users for a tariff of 100 euros per hour, giving labor costs of 3000 euros per month. Then some costs for hardware and software could be somewhere around 1000 euros per month (considering, e.g., write-off of advanced computing servers or subscription for cloud computing and storage). Now the benefit falls greatly to the advantage of the smart routing application. However, initial costs for development of this service (to make it mature and ready for large-scale use) are also substantial and can, for example, be estimated around 100 thousand euros. A business model could be used to ask a small amount from the users for this service (let us say 2 euros for downloading the app and 2 euros per month for using it), in order to gain back the initial development costs quickly.

10. Conclusions

The results of this study give insight into the size of improvement that is possible with individual in-car routing advice, for different traffic situations, and which quality of the input traffic data is needed to achieve a certain degree of improvement. This improvement is expressed with traffic performance indicators as total travel time and total delay.

For a penetration rate of 10% on the average day, when perfect traffic data and traffic information would be available, the total improvement over the whole day is 0.8% of the total delay in the network, which means a saving of about 15 million euros on a yearly basis. For 90% penetration rate, 7.0% of the total delay in the network would be saved, which is approximately 130 million euros on a yearly basis.

Though the previous results are based on perfect information, this study furthermore shows that even with low data availability and low speed accuracies the delays in the network are improved. For 90% penetration rate, on a regular day the improvement in saving delay hours is about 7%; on a day with incidents this increases to 8.6%. When only looking to the morning peak, the difference in effect for a normal day or a day with incidents is larger: from 4% on a regular day to 14% on a day with incidents.

When the link speed accuracy is less than 40%, the effect of the smart routing advice becomes negative, because it is based on the wrong information.

Also a relation has been made between the FCD penetration rate and the quality aspects of the data for the smart routing: with a penetration rate of at least 10%, the speed error based on small individual location errors from GPS stays below 6%. Therefore the FCD penetration rate has hardly any influence on the delay improvement from the smart routing. For a penetration rate of 1% this could not be estimated because of very little data. For the link availability, an FCD penetration rate of 1% leads to link availability of 5%. For a penetration rate of 10% or higher, the link availability is so high that again the FCD penetration rate has not much influence.

A coarse cost-benefit estimation has been done based on rough estimates of various costs (fixed and variable). This shows that the necessary communication costs for a mobile data connection (subscription with data for the users) do not weigh against the benefit in value of time per day, however,

since most people already have a mobile phone subscription with data connection, the extra costs for this service are zero. Maintenance costs are very low compared to the benefits. Initial development costs are estimated substantial but can be gained back quickly by asking a very small amount from the users for this service.

11. Recommendations for Further Research

Though in this case study a good attempt has been made for the estimation of the potential effects of a large-scale smart routing service, some issues could be studied in more detail in order to improve the reliability and applicability of the results. Additional questions and steps for further research are as follows:

(i) In order to get more reliable results on a yearly level, scale up the results by using more days with different traffic patterns and investigate how many abnormal days with more congestion than average usually occur throughout a year.

(ii) Use a route choice model that takes into account overlap of routes, calibrate with real data.

(iii) Calculate personal benefits (in addition to network effects)

(iv) Improve the routing advice taking into account capacity constraints: when the route with the current shortest travel time almost exceeds capacity, advise the 2nd (or 3rd) best route. This can, for example, be done by performing an equilibrium traffic assignment based on actual traffic measurements.

(v) Improve the routing advice by updating the advice while driving, using actual traffic measurements. This was already implemented in practice in the Practical Trial Amsterdam [2].

Since this study shows that the penetration rate and FCD data accuracy are key elements in relation to the delay improvement, strategies that could increase the penetration rate of such in-car services and possibilities to improve data quality from FCD are relevant topics for further research as well. For example, additional research can be done on the accuracy of the localization: does localization based on gsm signal strength provide sufficient accuracy or is GPS accuracy needed? In the first case penetration rates are expected to be higher since not everybody has or uses GPS localization all the time on his mobile phone.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

joint collaboration between TNO, Delft University of Technology, and Rijkswaterstaat, highway agency of the Dutch Ministry of Infrastructure and the Environment.

References

[1] T. Djukic, I. Wilmink, E. Jonkers, M. Snelder, and B. van Arem, "Exploratory analysis of traveller's compliance with smartphone personal route advice: a field trial Amsterdam," in *Proceedings of the 95th Annual Meeting of the Transportation Research Board*, 2016.

[2] I. Wilmink, E. Jonkers, M. Snelder, and G. Klunder, "Evaluation results of the Amsterdam, Netherlands, practical trial with in-car travel and route advice," *Transportation Research Record*, vol. 2621, pp. 38–45, 2017.

[3] G. A. Klunder, H. Taale, L. Kester, and S. Hoogendoorn, "The effect of inaccurate traffic data for ramp metering: Comparing loop detectors and cameras using information utility," in *Proceedings of the 19th World Congress of the International Federation of Automatic Control*, pp. 5075–5082, Cape Town, South Africa, 2014.

[4] F. Viti, S. P. Hoogendoorn, L. H. Immers, C. M. J. Tampère, and S. H. Lanser, "National data warehouse how the Netherlands is creating a reliable, widespread, accessible data bank for traffic information, monitoring, and road network control," *Transportation Research Record*, no. 2049, pp. 176–185, 2008.

[5] V. Sethi and F. S. Koppelman, "Incorporating complex substitution patterns with distance and variance scaling in long distance travel choice models," in *Travel Behavior Research, the Leading Edge*, D. A. Hensher and J. King, Eds., Pergamon Press, Oxford, UK, 2001.

[6] S. C. Calvert, M. Snelder, T. Bakri, B. Heijligers, and V. L. Knoop, "Real-time travel time prediction framework for departure time and route advice," *Transportation Research Record*, vol. 2490, pp. 56–64, 2015.

[7] H. Taale and S. P. Hoogendoorn, "Network-Wide Traffic Management with Integrated Anticipatory Control," in *Proceedings of the 91st Annual Meeting of the Transportation Research Board*, Washington D.C., USA, 2012.

[8] I. Wilmink, E. Jonkers, A. Jöbsis et al., "Eindrapport Evaluatie Praktijkproef Amsterdam," in *CAR–Perceel Regulier Verkeer*, TNO 2016 R10044, TNO, Delft, Netherlands, 2016.

[9] M. Modsching, R. Kramer, and K. ten Hagen, "Field trial on GPS Accuracy in a medium size city: The influence of built-up," in *Proceedings of the 3rd workshop on positioning, navigation and communication (WPNC'06)*, 2006.

[10] T. Chalko, "Estimating accuracy of GPS doppler speed measurement using speed dilution of precision (SDOP) parameter," *NU Journal of Discovery*, vol. 7, pp. 4–9, 2011.

[11] C.-H. Wen and F. S. Koppelman, "The generalized nested logit model," *Transportation Research Part B: Methodological*, vol. 35, no. 7, pp. 627–641, 2001.

[12] H. Taale, *Integrated anticipatory control of road networks: A game-theoretical approach [Ph.D. thesis]*, Delft University of Technology, 2008.

[13] E. Cascetta, A. Nuzzolar, F. Russo, and A. Vietta, "A modified logit route choice model overcoming path overlapping problems," in *Proceedings of the the 13th International Symposium on Transportion and Traffic Theory*, pp. 697–711, Lyon, France, 1996.

[14] S. Hoogendoorn-Lanser, R. van Nes, and P. H. L. Bovy, "Path size and overlap in multimodal transport networks," in *Proceedings of the 16th International Symposium on Transportation and Traffic Theory*, pp. 63–84, College Park, Md, USA, 2005.

[15] S. C. Dafermos and F. T. Sparrow, "The traffic assignment problem for a general network," *Journal of Research of the National Bureau of Standards*, vol. 73B, pp. 91–118, 1969.

[16] P. Warffemius, "De maatschappelijke waarde van kortere en betrouwbaardere reistijden. Kennisinstituut voor Mobiliteitsbeleid, 2013".