Real-time travel time prediction framework for departure time and route advice

Calvert, SC; Snelder, M; Bakri, T; Heijligers, B; Knoop, VL

Publication date
2015

Document Version
Accepted author manuscript

Published in
Proceedings of the Transportation Research Board 94th annual meeting

Citation (APA)
Calvert, SC., Snelder, M., Bakri, T., Heijligers, B., & Knoop, VL. (2015). Real-time travel time prediction framework for departure time and route advice. In s.n. (Ed.), *Proceedings of the Transportation Research Board 94th annual meeting* (pp. 1-15). Transportation Research Board (TRB).

Paper no: 15-3916: Submitted to the 94th Annual Meeting of the Transportation Research Board
11-01-2015-15-01-2015, Washington, United States

Real-Time Travel Time Prediction Framework for Departure Time and Route Advice

Simeon C. Calvert (Corresponding author)
TNO
Van Mourik Broekmanweg 6, P.O. Box 49,
2600 AA Delft, The Netherlands
+31 88 866 3314
simeon.calvert@tno.nl

Maaike Snelder
TNO
maaike.snelder@tno.nl

Taoufik Bakri
TNO
taoufik.bakri@tno.nl

Bjorn Heijligers
TNO
bjorn.heijligers@tno.nl

Victor L. Knoop
Delft University of Technology
Department of Transport & Planning
TRAIL research school
TrafficQuest
Stevinweg 1, 2628 CN Delft, The Netherlands
+31 15 278 8413
v.l.knoop@tudelft.nl

Submitted: July 31, 2014

Total words: (Words: 6700; Figures: 4x250) = 7700

ABSTRACT

Heavily utilized urban networks remain a challenge for travel time prediction, as traffic flow is rarely homogeneous and is also subject to a wide variety of disturbances. Various models, both based on traffic flow theory and data-driven, have been developed to predict traffic flow and travel times. Many of these perform well under set conditions. However few perform well under all or even most urban traffic conditions. As part of the Amsterdam Practical Trial, a comprehensive FOT for traffic management, a real time travel time prediction framework has been developed that makes use of an ensemble of traffic modelling techniques to be able to predict travel times with great accuracy for arterial roads as well urban roads. The various models in the framework comprise of both traffic theoretical models as well as data-driven approaches, making use of some of the largest real-time traffic datasets currently available to traffic engineers to limit errors to less than 20% for any time of day or week. The impending practical implementation of the framework sets it at the forefront of practical real-time implementation of urban travel time prediction.

1. INTRODUCTION

It is commonplace for road networks in urbanized areas to be heavily utilized throughout the world. This obviously allows travel and therefore enables mobility of inhabitants, however overuse of the urban road networks also leads to wide spread congestion, emission population and noise problems. The focus of this paper is on pre-trip and en-route dynamic travel information and advice. Moreover, the focus is on the travel time predictions that are required to give a high quality advice. Currently most internet services and applications base their mode, departure time and route choice information and advice on historical and instantaneous travel times. However, improved travel advice can be given if travel time predictions are taken into account *(1, 2)*.

Various models, both based on traffic flow theory and data-driven, have been developed to predict traffic flow and travel times. They can be classified into static and dynamic models, into microscopic, mesoscopic and macroscopic models, into offline and real-time models, into data-driven and model-driven models and into deterministic and stochastic models. Also, within these classes different methods may be distinguished. As each of these models has its strengths and weaknesses there is not one method that outperforms the others under all conditions. Different forecast horizons, different locations, different road types, different times of day and different situational variables like incidents, events and road works demand different prediction models and approaches. Therefore, this paper aims to introduce a dynamic real time travel time prediction framework in which an ensemble of data-driven and simulation based traffic models are run in parallel and the best prediction is selected depending on the conditions. Furthermore, a new short term prediction model is introduced for incident situations on the motorways.

The proposed framework is applied in the Amsterdam Practical Trial which aims to test in-car information services on a large scale in the Amsterdam region in the Netherlands. The aim is to bring dynamic and personalized traffic information into the vehicle, thereby improving the reliability of travel times in and the accessibility of the Amsterdam region. These dynamic travel times further improve on many current approaches as forecasted predictions, rather than giving instantaneous travel times as a prediction.

In section 2 of this paper a general overview is given of model types that can be used for short term predictions. Section 3 presents the model framework, with the prediction models that are applied in the Amsterdam Practical Trial presented in section 4. Section 5 describes how the framework and the prediction models are applied in the Amsterdam Practical Trial. A quality indication of the prediction models is also presented in this section. In the last section the main conclusion and future research directions are presented.

2. EXISTING METHODS

There are many types of traffic prediction models in existence. This paper focusses on methods for travel time prediction, which can be performed by traffic prediction based on traffic properties (such as simulation models), or by numerical – mainly data-driven – techniques to predict travel times. These two methods will be discussed in the sections below.

Simulation methods

A first approach for describing traffic operations are microscopic simulation models. These models describe individual vehicles and the behavior of drivers. Commercial packages are available, of which Vissim *(3)*, Aimsun *(4)* and Paramics (*5*) are among the most used packages. In this paper, we will not go into detail regarding differences in these packages, but comment on the category of the description instead.

Microscopic models generally have a longitudinal module and a lateral module describing the vehicle movements. The longitudinal module describes car-following behavior in case a driver is following another vehicle and a free flow speed choice in case it is not. The lateral module describes a lane choice, for instance for overtaking, merging or for drivers which want to pre-select a lane because they need to take a turn or exit. These types of models are adapted to suit all kinds of environments. They can be applied on the freeway and highway, but also in urban environments. Some models include cyclists and pedestrians as well. The output from the models can be very detailed, up to the position of all road users for all moments in time. Visually, this approach is very appealing, since the output can usually be visualized in a way which is very attractive and looks realistic. There are also disadvantages. First, the behavioral models are usually not calibrated for the situation where the model needs to be applied. That could be considered a task for the model user; however, in some cases this is impossible because there are unknown measures or the infrastructure is not yet built. Moreover, it is not clear whether realistic individual behavior will lead to realistic collective behavior, and the properties of this collective behavior will in the end lead to congestion, and hence determine the travel times. Furthermore these models are computationally demanding, especially for larger areas. Note that for larger network areas it is often required to predict a longer time ahead *(6)*. A larger area means also a larger number of road users, and therefore also larger computation times, which are the core units of the microscopic simulation approaches. The computation speed might become less than real time, which hampers such models as an option to use for travel time prediction.

Travel time prediction can also be performed with *macroscopic models*. In this case the traffic flow is described at an aggregated level using traffic flow theory. This uses the relationship between average speed and density in case first order models are used. For higher order models there is a form of hysteresis built in *(7)*. The Cell Transmission Model *(8, 9)* is a well-known model describing traffic dynamics for road split into cells. The Link Transmission Model (*10*) does not need the separation into cells, and can therefore run quicker. The advantage of these models is that all knowledge of collective patterns can be brought into the model, but for unknown situations the collective patterns are more difficult to predict than the reaction of individual road users. Moreover, the models work very well for uninterrupted flow facilities (freeways) but in the urban area their use is limited. A good node model *(11)* is essential to apply the model in an urban area.

An area of macroscopic modelling that has shown recent promise for fast and efficient calculation is that of marginal traffic modelling. By only simulating the marginal difference in traffic flow in comparison to a base run, repetitive network loading with a full dynamic macroscopic model is not required. This approach has been shown to especially be effective when a large number of predictions are to be made in which the traffic characteristics only change by a small amount (*12*-*14*). For real-time traffic prediction, such models are effective due to the short calculation times and the relatively marginal changes in the traffic states from one minute to a next.

Data driven methods for travel time predictions

Another category for predicting travel times is to not get into the theory of driving, but instead use data and computer learning to derive patterns from traffic flow to produce a travel time prediction. In general, different models can be used to predict travel times. There are set situations which are used to train the models. Then, measurements are taken for specific days and the model is used to predict the travel time. The more complex the models are, the more complex patterns they can capture, but there is also a risk of over fitting, which means that a model has more parameters than needed and the model adapts to the stochastic fluctuations of the learning set, instead of only capturing the underlying patterns. The simplest models are regression models. These can be combined and for different cases the best model can be chosen *(15)*. The principle of having several models to choose from is increasingly used, and is called an ensemble. The idea behind an ensemble is that several models are run, and that the user continues with a combination of the outcomes of these models. In fact, the model which is chosen to continue with might depend on the outcome. A very simple scheme for a travel time prediction would, for instance, be to take the average travel time of all model outcomes, except for the two most extreme values. More advanced systems would rate the reliability of the outcome to the outcome itself: for instance, if the bandwidth of travel times is high, the quality of the prediction might be low, hence none of the models are selected.

Also neural networks are a possibility to predict travel times *(16)*. The combination between the inputs and the outputs is given by one or more intermediate calculation steps in a hidden layer. The extent to which the inputs activate a hidden layer and to which these hidden layers process their values to the output is determined in the learning process.

Real-time prediction methods

For on-line traffic control, real time traffic models are required. This means that the models apply input from traffic sensors and use that to make a prediction. That prediction, in turn, is used as input for traffic control. It is therefore essential that the time required to make a traffic prediction is quicker than real-time. If this is not the case, a prediction will already be outdated by the time it is made and is therefore useless. In that case a naïve prediction, taking the current instantaneous travel time, would be a better prediction. Moreover, this information is not useful for controlling traffic. On-line means that the traffic model is fed with real world data and the traffic model is hence connected, this opposed to models which are only fed for instance with predetermined OD matrices. It is therefore essential for these models that they are fast. Moreover, they should be robust for failing, or erroneous input. Detectors give erroneous data, and although many systems are developed to improve the data (e.g. *(17)*), still the traffic state estimation might be incorrect. A good model is able to avoid output which very strongly fluctuates based on single missing or erroneous measurements. For a comprehensive overview of existing short-term prediction models, we recommend the recent work by Vlahogianni, Karlaftis (*18*).

3. FRAMEWORK

In this contribution we present a model framework that comprises of multiple sub-models to predict door-to-door travel-times for an entire network in real-time. The architecture that is used to give the advice is shown in Figure 1. On the left side of the figure the different data sources that have been used are shown. The data sources contain historical and real-time data. A map matching algorithm is used to match the data to a map. Based on this fused data the prediction models are run. The results of the prediction models are combined by the Hypothesis Manager (HM). In the following section it is explained how the HM combines the results of the different local, arterial and incident predictors. For the main arterials, the results of the an Arterial Network Predictor (ANP) are always used unless this predictor is overruled by the Marginal Model for Incidents, which is applied in case of unexpected incidents. The combined predictions for the Local Network Predictor, which is applied on lower level roads, are used for the other road types. The before mentioned steps can only be performed for the roads for which data is available. In order to be able to give a smart routing advice, also realistic travel times and travel time predictions need to be available for the other roads. Therefore, a combination of gap filling algorithms is implemented, which is described in the following section. The result is that each 5 minutes new short term predictions (three hours ahead) for all roads in the network become available for a smart routing algorithm which is used to give different advices. The map matching, data fusion, different advice strategies and smart routing are not subject of this paper. In the following section, each of the prediction models are explained in greater detail.

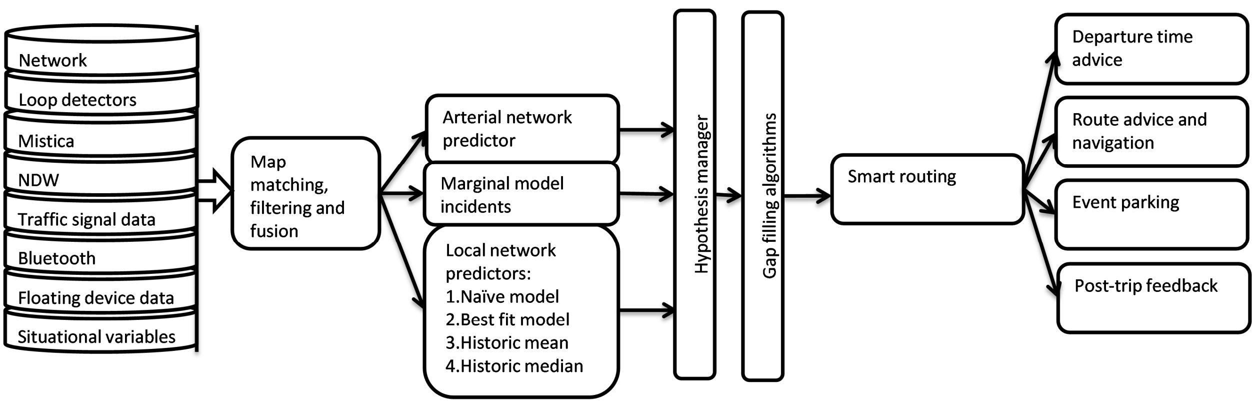

Figure 1 Architecture Practical Trial Amsterdam

4. APPLIED METHODS

The model framework is developed initially for use in a large scale field-operation-test (FOT) known as the Amsterdam Practical Trail. The various sub-models in the framework interconnect to give the overall predictive framework. For non-incident situations a combination of data driven models is chosen. An important reason is that the computation time of pre-trained data driven models is generally lower than simulation based methods, which is imperative for real-time application. In case of incidents, data-driven models are less suited because by definition incidents are exceptional situations for which historical data does not offer a good prediction basis. Here the three developed methods which are applied as sub-models in the framework are expanded on. Firstly the prediction methods applied as a base for all road types and the specific local network predictor are described. Thereafter the Arterial Network Predictor and then the method for predicting under incident conditions are given.

Hypothesis Manager and Local Network Predictor

Many data-driven approaches have been developed to predict travel times based on historical and real-time data. Some of these exclusively use instantaneous travel times and are limited for flexibility. Many of these approaches are deliberately keyed towards certain patterns in traffic and therefore perform well under specific conditions, but rarely under all traffic conditions. Approaches that can predict under nearly all conditions may demand a greater deal of complexity, which negatively influences practical real-time application. It is obvious that an aggregation of multiple approaches which complement each other and make use of each approaches strength should lead to a better prediction result. For this reason, and for easy and robust implementation, the Local Network Predictor (LNP) is developed in combination with a Hypothesis Manager (HM). Within the LNP an unlimited number of prediction models may be plugged in each with its own prediction strengths. The Hypothesis Manager is developed to process the predictions from the LNP and produce an overall aggregated travel time prediction based on the best predictions from each individual model in the LNP. Before looking at the HM in more detail, we will first expand on the LNP and its initial models.

The Local Network Predictor, or LNP, exists of four prediction models in its current form, but is designed for any arbitrary number of models. As the application of the overall framework demands fast and robust real-time prediction, the applied models in the LNP are deliberately kept simple. The applied models are: 1) Naïve model; 2) Best fit model; 3) Historic mean; 4) Historic median

The *naïve* prediction model is the most basic model and returns a prediction identical to the flow and speeds recorded in the previous minute. The *best-fit* model makes a comparison between the past T minutes, set initially to 60 minutes and the previous D days of traffic flow, set initially to 100 days, and selects the day which matches best as the prediction for the coming period. Both the *historic mean* and *historic median* predictors consider a period of W weeks, with the default set to 7 weeks, in the past and take the mean and median realizations respectively as the prediction. For each of the predictors, a distinction is made between four prediction variables: road category, day of the week, time of day, and the prediction horizon.

The Hypothesis Manager (HM) is developed to evaluate the quality of predictions from the individual models from the LNP and concatenate the results from the combinations of the various prediction variables. The HM exists of three main parts, namely the *evaluator, the HM predictor, and the gap-filler* (see Figure 2a). The predictions from each prediction method are evaluated differentiating between:

1. Road category
   Three categories (A, B or C) are defined based on the maximum speed limit: High, medium and low level roads based on a nominal maximum speed limit of >=100 kph, 70-80 kph, and <=50 kph.
2. Day of the week
   Each of the seven days of the week individually
3. Time of day
   In aggregated blocks of 5-minutes
4. Prediction horizon

From 0 minutes (current state) for 5 minutes intervals up to 3 hours in advance.

For each combination of the prediction variables, a prediction is made by each prediction method in the LNP. Each prediction is evaluated by the HM predictor against training data using the Mean Absolute Error as error measure:

$$E = \frac{\sum_{l=1}^{N} |P_l - R_l|}{\sum_{l=1}^{N} R_l} \tag{1}$$

Here P_l is the predicted travel time on link$l$, R_l is the realized travel time of the same link$l$, and N is the total number of links on the network. The model is initially trained for the year 2013 using extensive data collected in the databases of the National Data Warehouse for Traffic Information (NDW) in The Netherlands. This database comprises (real-time) traffic counts from the majority of the Dutch road network from double induction loops, floating car data, camera system, etc. (*19*).

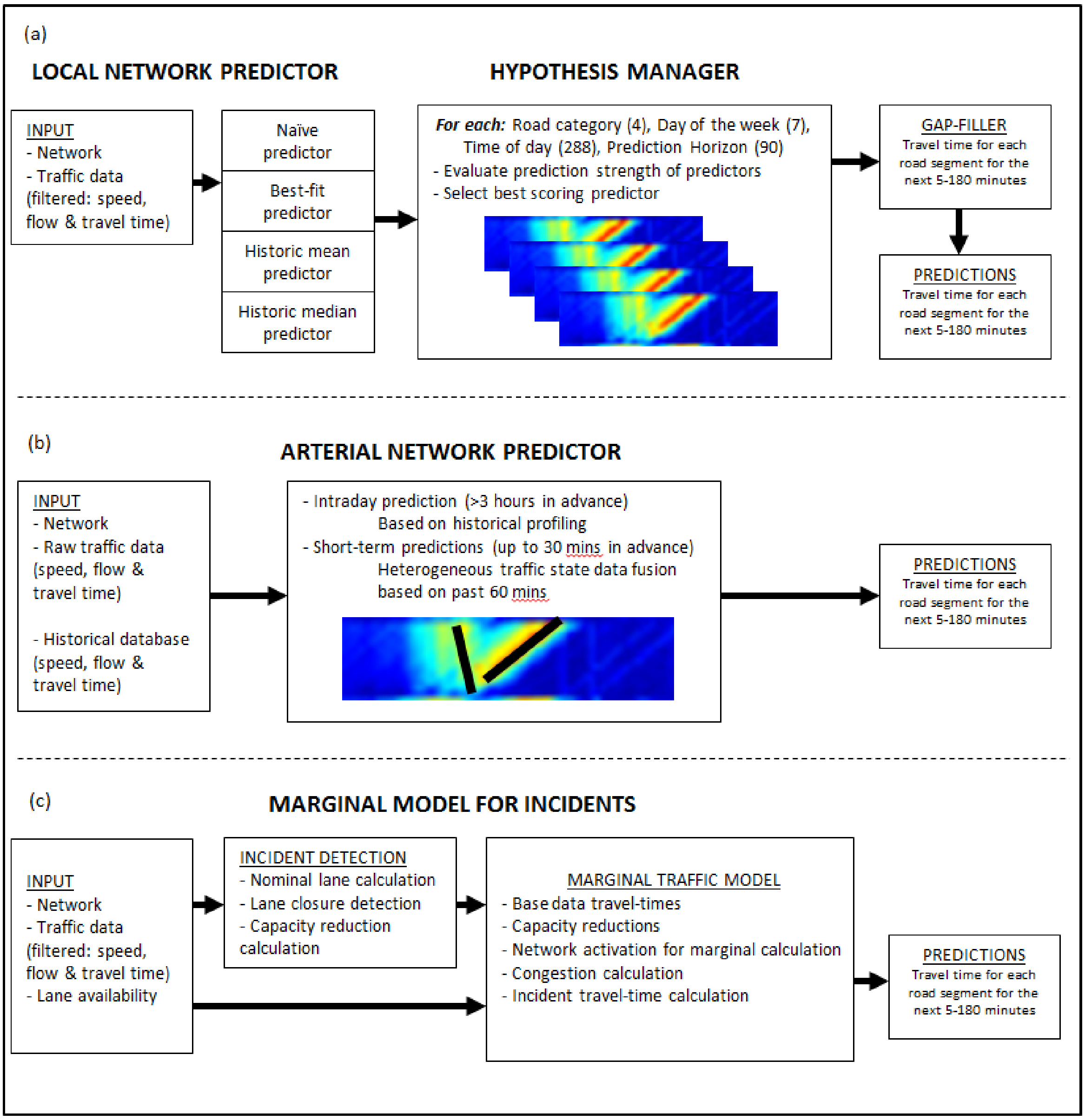

Figure 2: (a) Local Network Predictor and Hypothesis Manager (b) Arterial Network Predictor (c) Marginal Model for Incidents

Although a vast area on both the local and arterial networks is covered in the NDW data, gaps remain on the routes and roads where no travel time predictions or other data are available. This is also the case for locations with data of a low quality, which is filtered out in advance. To bridge the gaps for these network locations, a *gap-filler* is applied for the regions in which precise travel time predictions are made. It is necessary to fill these gaps, as otherwise correct prediction cannot be made along these road segments. The gap-filler is based on the principle of propagation and interpolation of speed patterns along a large set of predefined routes through the considered network. Some 1000 partially overlapping routes are defined in the considered network area (greater Amsterdam) (see Figure 3) making use of Dijkstra's algorithm for the shortest travel time. Multiple weighted interpolation of traffic speeds through the unknown parts of the network is performed making use of the known speeds on links of the routes that cross the links which are unmeasured. The weighted interpolation considers up to 10 links along each overlapping route for a distance of 1.5 kilometers. A weight is assigned to each speed observation depending on the road type and the distance from the considered road section. This approach allows the predominant traffic sates from the same corridors to be translated to the considered road sections, such that a good prediction can be made. Additional back-up predictions are made based on the network speeds within a 1.5 kilometer radius of the link in case insufficient routes or insufficient speed observations are available.

Arterial Network Predictor

The *Arterial Network Predictor* (ANP) is applied for travel time predictions on main arterial routes for which a good quality of traffic data is consistently available in both time and space (Figure 2b). Motorways in The greater Amsterdam Area are nearly entirely fitted with double induction loops as part of the Monitoring Casco (MONICA) system and are therefore available for predictions using the ANP. As part of a personalized route advice, the ANP predicts travel times on two levels: those with a long term horizon (interday) and those with a short term horizon (intraday up to 30 minutes to 3 hours in advance).

The *Long term prediction* is applied for predictions made one or more days in advance to a trip being made. For this reason the prediction does not take the current traffic states into account, rather the expected states based on historic traffic patterns. This is performed by taking the median value of the four previous days matching the same day-of-the-week, including filtering for irregular days such as holidays.

The *Short term prediction* is performed in real-time using *heterogeneous traffic state data fusion* as originally described by *(20-22)* and later adapted for fast and efficient application by Schreiter et al *(23, 24)*. The method makes use of spatiotemporal patterns in traffic flow combined with kinematic waves defined in traffic flow theory. Predictions are made through propagation of these spatiotemporal traffic patterns into the future to give a future traffic state (speeds and flows) from which travel times can easily be derived. The past 60 minutes of traffic data is applied as input for the method, which makes predictions for the following 30 minutes on each road segment. Predictions more than 30 minutes in advance up to a prediction horizon of 3 hours are made through a combination of the long and short term predictors. The reason behind this is that the traffic propagated spatiotemporal traffic patterns in the short term prediction are subject to change in time, which cannot be accurately predicted with certainty beyond the initial 30 minute period. The method is applied as of the shelf and therefore we refer to the relevant papers for details and equations (*23*).

Marginal Model for Incidents

While data-driven models can boast success in pattern recognition and prediction, predicting random and spontaneous events remains out of their scope. Such events, such as incidents, therefore require an alternative approach. The *Marginal Model for Incidents (MMI)* is a real-time marginal model, which is fed with live traffic data and includes an incident detection algorithm. The flow diagram in Figure 2c shows each part of the model.

Marginal traffic models are models that generally make use of the outcome of a base model run and only update areas of a network which have been significantly altered due to changes either in the traffic flow or in the network characteristics. This makes such models extremely fast and efficient, while remaining sufficiently accurate (*14*). More information on marginal traffic models in general can be found in *(12-14).* The parts of the network deemed necessary to update are known as the activated network. The MMI is a hybrid marginal model in the sense that it is both data-driven and theory-driven. Data of the current traffic states and the characteristics of the network are applied as the 'base model', while the modelled prediction for the activated parts of the network is performed using traffic theory. Activation of the network is based on the presence of congestion, which negatively influences speed and therefore travel time. From the live traffic data an estimate is made of the traffic state and available unused capacity. Road sections with congestion as a consequence of an incident are added to the activated network. Upon occurrence of an incident, the available capacity is reduced and a calculation is performed how quickly the remaining capacity is filled. The capacity reduction is determined by the number of closed and available lanes, together with empirical data from previous incidents of similar types on similar road sections. If capacity is exceeded, upstream links are activated and congestion is propagated in time using kinematic wave theory to calculate the speed of the shock wave. The upstream distance that the shockwave lasts is calculated using the difference between the inflow at the end of the queue and outflow after the incident, and the available capacity of upstream links. This process continues until the upstream end of congestion is reached. Using the time congestion takes to reach upstream links and using knowledge of the severity of congestion allows one to make a prediction of future travel times for different prediction horizons. This process is repeated every minute using the flow data from the time just before the start of the incident and updated for the live traffic conditions. This updating is performed by applying a feedback loop that compares the estimations of the MMI with the real travel times from data and applies a correction factor to the capacity reduction in the following minute. This feedback correction is continuously applied to allow an increasingly accurate prediction of the travel times due to the incident. It also allows for indirect correction of the presumed capacity reduction as the actual capacity reduction can only be initially estimated based on the number of closed lanes.

The detection of incidents is performed using two sources: primarily through automatic lane closure detection from the Dutch highways agency (Rijkswaterstaat). Secondly through the incident registration system. The lane availability data is available along with the live traffic speed and count data and gives the lane availability with a delay of less than one minute. Detection of closed lanes however does not necessarily mean an incident has occurred, because overflow, peak-period, and tidal lanes are often closed during the day when traffic is quieter. Therefore a lane availability detection algorithm is applied, which filters the number of available lanes for a specific day and is compared to the number of open at any given time. This process is updated daily to avoid the detection of roadworks as incidents, presuming that roadworks are present for an entire day, which is reasonable for The Netherlands. Using an empirical database of incident types, including the number of closed lanes and the type of road, an estimate is made of the initial reduction in capacity. The capacity reduction is later adjusted in the feedback loop as previously described. The secondary incident detection through the registration system is applied to give further information on the type and extent of an incident. It is applied as a back-up and supplementary system as there is often a delay in registration of at least five minutes and always exceeds the time required for the lane closure detection update.

5. CASE STUDY: AMSTERDAM PRACTICAL TRIAL

This section describes how a mixture of the above mentioned short term prediction models have been applied in the Amsterdam Practical Trial the Netherlands. The aim of the field trial is to bring dynamic and personalized traffic information and advice into the vehicle, thereby improving the reliability of travel times in and the accessibility of the Amsterdam region. Over 10.000 participants will be given a departure time advice, route and navigation advice, parking advice for major events and post trip feedback by the end of 2014 and in 2015 by means of an internet web service and a smartphone application.

Data sources

The network that is used is shown on the right side of Figure 3. The network contains about 116 thousand links and 68 thousand nodes. The network includes all the motorways in the Netherlands and has a higher level of detail in Amsterdam and surroundings as is shown on the left side. Especially the areas where events held have a high level of detail. The network does not contain all the roads in the Netherlands to limit the computation time and, more importantly, to focus the smart routing advice on the higher level roads. At most the first and last 1.5 kilometers of the navigation advice is given by standard navigation software.

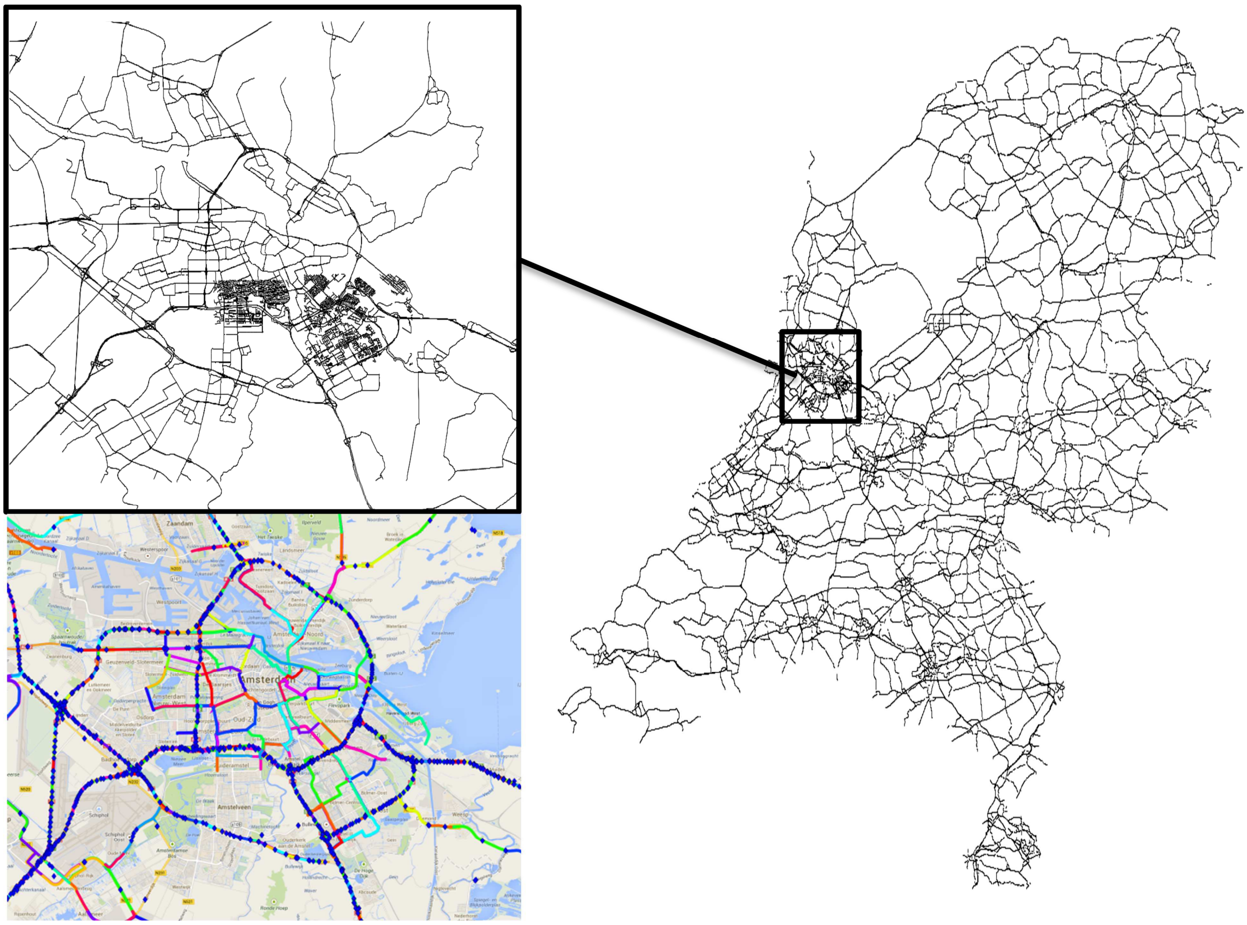

Figure 3 Road network Field Trial Amsterdam

For most of the motorway network historical and real time (1 minute delay on processing and up to a maximum of 3 minutes in the live feed) loop detector data is available. On average the distance between the loops is about 500 meter. For each minute, average vehicle speeds (km/h), flows (veh/min) and the lane closure status (Mistica) are stored. The National Data Warehouse for Traffic Information (NDW) provides historical and actual minute data about travel times and traffic volumes on the motorways, secondary roads and urban thoroughfares of the participating authorities. The roads for which the NDW provides data for Amsterdam and surroundings are shown on the bottom left side of Figure 3 (the colored links). These are the most important primary and secondary roads of Amsterdam. Travel time information is available on all colored links in Figure 3. On the blue dotted links (mainly motorways) also traffic volumes are available. Since the NDW data on the motorways is less detailed than the loop detector data with lane closures, the motorway prediction model and the marginal model for incidents use the loop detector data and the data for lane closures. The local network predictors use the NDW-data. In the near future additional data sources will be added for the local prediction models. Additional Bluetooth sensors will be installed in the event areas and floating device data from the application that is to be developed and from other applications will be used to make improved predictions. As stated above some gap filling algorithms have been developed to estimate travel times for roads where there is no data available or the quality of the data is not sufficient. Since the new data sources should improve the data quality and reduce the gaps, the role of the gap filling algorithms is expected to become less important in the future. Finally, situational data like rain data, road closures and the time of day and the day of the week is used to improve the quality of the predictions.

Quality of the prediction algorithms

In the previous sections it is explained how the different prediction models and the hypothesis manager work. In this section, a quality indication is given of the LNP. The other prediction models remain in the initial implementation phase and have yet to be evaluated. Figure 4a-c summarizes the quality of the four lower level predictors for the most important secondary roads (category B roads) for different prediction time horizons (y-axis) and different days of the week (x-axis). The shown quantity is the error which is the error measure as shown in Equation 1. The historic mean and median were almost identical and therefore only one is shown. Within each day, each 5 minute period of the entire day is plotted (left is 0.00 h and right is 23.55 h). The prediction results are compared with the realizations of the travel times. The results are averaged for each day of the week and all the links to come to an average prediction error. The results show that the naïve predictor performs well for predictions horizons up to 30 minutes (Figure 4a). This can be logically explained by the fact that it is likely that in a few minutes the traffic situation will not change much. For Saturdays and Sundays, the naïve predictor and the other predictors also perform well because most of the time on Saturdays and Sundays the network is in a free flow state which is easy to predict. For the longer term predictions the historic median predictor is shown to perform best (Figure 4c). In practice this predictor appears to almost always outperform the historic mean predictor. The best fit predictor performs best at the transition phase between peak and off-peak period and between the off-peak and peak period (Figure 4b). Eventually, the Hypothesis Manager determines which predictor is used on different days, different time of days, different prediction horizons and different road category. The most accurate, and therefore applied, predictor is shown in Figure 4d.

Figure 4e-g shows the results for relative error for each road type. The figures show that the predictions for the motorways (category A) are the most accurate and have on average a relative error of up to 20% depending on the prediction horizon and time of day. The largest errors can be found during the morning peak on Tuesdays. The prediction quality for category B roads is a bit lower than for A roads, but on average remains below 25%. For category C roads the errors are again slightly larger, especially during the off peak analyses. A further analysis indicated that this was caused by poor data quality on a number of roads. The quality of this data is currently being improved.

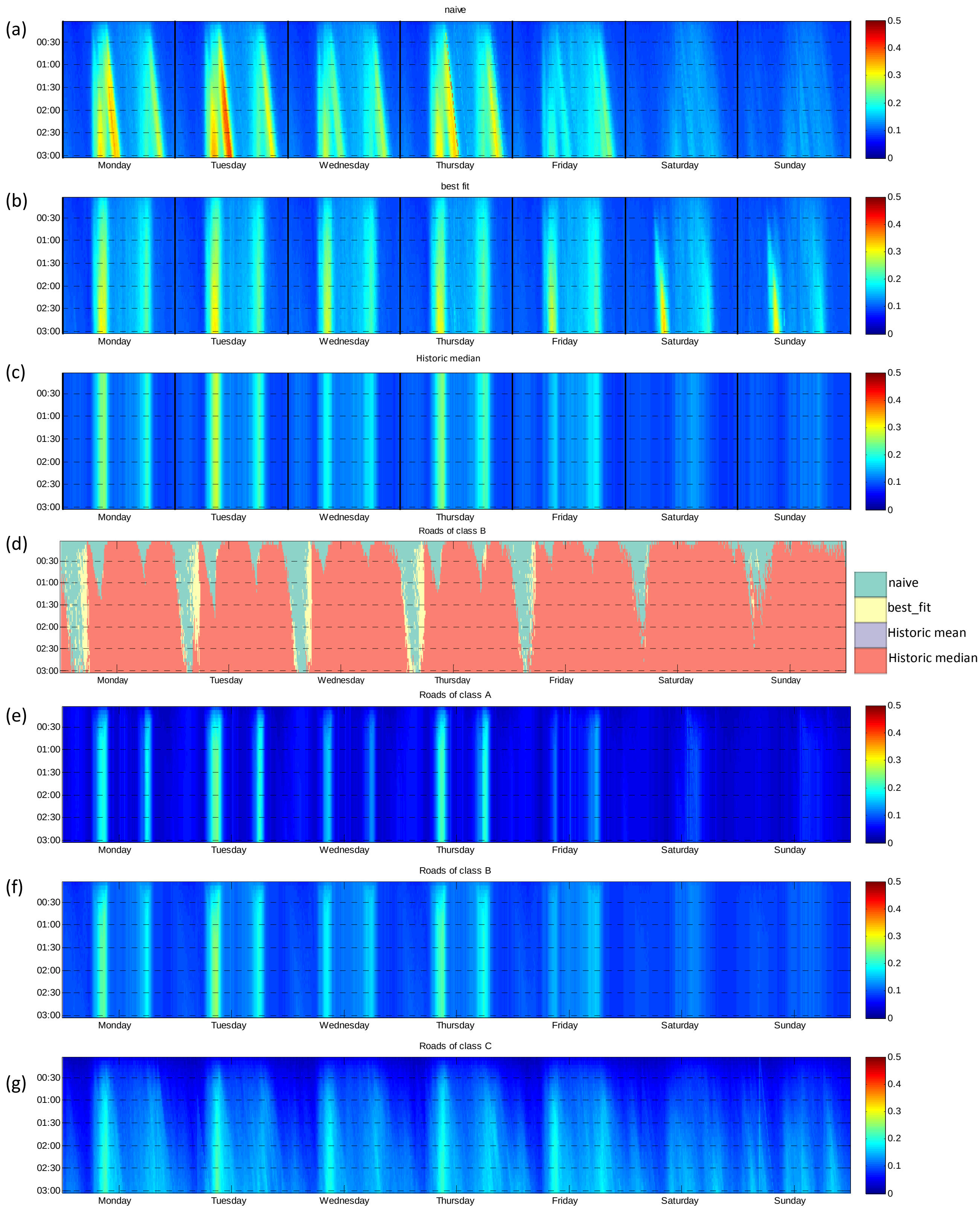

Figure 4: (a-c) Quality of the local predictors (naïve, best-fit & historic median) for cat B roads. (d) Applied model for time-of-day and day-of-week. (e-g) Overall prediction error for local predictors per road category. ***The axes show: Horizontal: time during a week. Vertical: prediction horizon.***

6. CONCLUSIONS AND FUTURE WORK

In this contribution a framework for real time predictions has been introduced in which different prediction models are run in parallel. It was advocated that it is a good choice to combine different models since different models perform well under different forecast horizons and for different locations, different road types, different times of day and different situational variables like incidents. This is demonstrated in the test case for the Amsterdam Practical Trial in which four data driven prediction models were selected and applied: a naïve prediction model, a best fit prediction model, a historic mean prediction model and a historic median prediction model. It was shown that for this specific case the naïve method, which assumes that the network state does not change, performs best for short prediction horizons (up to 15 minutes) and for uncongested traffic in the peak periods. For the longer term the historic median prediction model is especially suited and outperforms the historic mean prediction model. For the period just before and after the peak period the best fit prediction model performs best.

The initial results of the Amsterdam Practical Trial showed that these simple prediction models already perform quite well since the average absolute relative error of the combined models is in most cases below 20%. The advantage of the framework is that it is very easy to add more advanced prediction models to the framework due to the Hypothesis Manager, that was introduced as well in this paper, which can combine the results of many different prediction models into a single prediction.

The predominant choice for data-driven models over simulation based models was made because data-driven models avoid many calibration issues and can operate with shorter calculation times in real-time application. Of course the parameters of data driven models need to be calibrated as well, however this can be performed offline and is much easier than real time OD-estimation or calibration of the road capacities. Furthermore, if data driven models are used the actual traffic state is correct by definition (as long as the data is correct), whereas simulation models already deviate from the actual traffic state. Furthermore, the computational complexity of data-driven models is lower compared to simulation based models. A prediction model for arterial roads and incidents on arterial roads was also introduced. Opposed to the other prediction models the incident model is not a fully data-driven model due to the inability of data driven models to predict unknown incidents, but rather a hybrid marginal-data model. The model does however use actual data and is calibrated in real time. Extensive results of these models in practice will be discussed in future work.

Finally, the application of the prediction models in the Amsterdam Practical Trial has shown that the proposed framework allows predictions to be made on a large network (116 thousand links and 68 thousand nodes) within a short computation times with an update frequency of 5 minutes and with large data volumes.

Future work will focus on adding additional data sources like floating car, floating device and Bluetooth data to the framework. The framework is already designed in such a way that this can relatively easy be carried out as all the data is transformed to link travel times. Of course this does require additional data fusion. Furthermore, the quality of the prediction models will further be analyzed especially for the arterial network predictor and marginal model for incidents. Finally, the Hypothesis Manager will be improved in such a way that it becomes possible to switch between the model selection rules based on local actual and recent traffic conditions.

ACKNOWLEDGMENTS

This research is supported by the TrafficQuest, a collaboration between TNO, Delft University of Technology, and Rijkswaterstaat, highway agency of the Dutch Ministry of Infrastructure and the Environment. The research is further supported by the Amsterdam Practical Trial (APT) commissioned to TNO and ARS, by the Director-General of Rijkswaterstaat, part of the Dutch ministry of Infrastructure and the Environment.

The authors would also like to thank Isabel Wilmink, Ernst Jan van Ark, Jasper van Huis, Marco Duijnisveld and Bas van den Heuvel for their valuable input.

REFERENCES

1. Bellemans, T., B. De Schutter, and B. De Moor, *Models for traffic control*. JOURNAL A, 2002. 43(3/4): p. 13-22.
2. Hegyi, A., *Model predictive control for integrating traffic control measures*2004: Netherlands TRAIL Research School.
3. Fellendorf, M. and P. Vortisch. *Validation of the microscopic traffic flow model VISSIM in different real-world situations*. in *Transportation Research Board 80th Annual Meeting*. 2001.
4. Casas, J., et al., *Traffic simulation with aimsun*, in *Fundamentals of traffic simulation*2010, Springer. p. 173-232.
5. Cameron, G.D. and G.I. Duncan, *PARAMICS—Parallel microscopic simulation of road traffic*. The Journal of Supercomputing, 1996. 10(1): p. 25-53.
6. Knoop, V., et al., *Relationship between application scale and maximum time latency in Intelligent Transport Solutions.* Transportation Research Record: Journal of the Transportation Research Board, 2013. 2380(1): p. 1-9.
7. Laval, J.A., *Hysteresis in traffic flow revisited: an improved measurement method.* Transportation Research Part B: Methodological, 2011. 45(2): p. 385-391.
8. Daganzo, C.F., *The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory.* Transportation Research Part B: Methodological, 1994. 28(4): p. 269-287.
9. Daganzo, C.F., *The cell transmission model, part II: network traffic.* Transportation Research Part B: Methodological, 1995. 29(2): p. 79-93.
10. Yperman, I., *The Link Transmission Model for dynamic network loading.* Ph.D. Thesis, Katholieke Universiteit Leuven., 2007.
11. Tampère, C.M.J., et al., *A generic class of first order node models for dynamic macroscopic simulation of traffic flows.* Transportation Research Part B: Methodological, 2011. 45(1): p. 289-309.
12. Corthout, R., et al., *Improving the efficiency of repeated dynamic network loading through marginal simulation.* Transportation Research Part C: Emerging Technologies, 2014. 41: p. 90-109.
13. Corthout, R., et al. *Marginal dynamic network loading for large-scale simulation-based applications*. in *Transportation Research Board 90th Annual Meeting*. 2011.
14. Himpe, W. and C. Tampère, *Efficient Dynamic Network Loading Modeling: The fixed point link transmission model.* BIVEC-GIBET Transport Research Day, 2013: p. 95-98.
15. van Hinsbergen, C.P.I. and J.W. van Lint, *Bayesian combination of travel time prediction models.* Transportation Research Record: Journal of the Transportation Research Board, 2008. 2064(1): p. 73-80.
16. Van Lint, J., S. Hoogendoorn, and H.J. van Zuylen, *Accurate freeway travel time prediction with state-space neural networks under missing data.* Transportation Research Part C: Emerging Technologies, 2005. 13(5): p. 347-369.
17. Chen, C., et al., *Detecting errors and imputing missing data for single-loop surveillance systems.* Transportation Research Record: Journal of the Transportation Research Board, 2003. 1855(1): p. 160-167.
18. Vlahogianni, E.I., M.G. Karlaftis, and J.C. Golias, *Short-term traffic forecasting: Where we are and where we're going.* Transportation Research Part C: Emerging Technologies, 2014.

19. Viti, F., et al., *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: Journal of the Transportation Research Board, 2008. 2049(1): p. 176-185.
20. Treiber, M. and D. Helbing, *Reconstructing the spatio-temporal traffic dynamics from stationary detector data.* Cooper@ tive Tr@ nsport@ tion Dyn@ mics, 2002. 1(3): p. 3.1-3.21.
21. Treiber, M., A. Kesting, and R.E. Wilson, *Reconstructing the traffic state by fusion of heterogeneous data.* Computer-Aided Civil and Infrastructure Engineering, 2011. 26(6): p. 408-419.
22. Van Lint, J. and S.P. Hoogendoorn, *A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways.* Computer-Aided Civil and Infrastructure Engineering, 2010. 25(8): p. 596-612.
23. Schreiter, T., et al. *Two fast implementations of the Adaptive Smoothing Method used in highway traffic state estimation*. in *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*. 2010. IEEE.
24. Van Lint, J.W., *Empirical evaluation of new robust travel time estimation algorithms.* Transportation Research Record: Journal of the Transportation Research Board, 2010. 2160(1): p. 50-59.