MODEL – BASED SHORT TERM PREDICTOR OF TRAFFIC STATES

L.C.W. Suijs
Goudappel Coffeng, Traffic Management Department
L.J.J. Wismans
University of Twente / DAT Mobility
L. Brederode
DAT Mobility

1 ABSTRACT

The quality and availability of traffic data has been significantly improved in the last few years. More and more, loop detector data and floating car data becomes available (real time). This offers opportunities for new tools for traffic flow analysis and prediction relevant for operational traffic management services. Being able to detect and predict incidents, e.g. queues, accidents and car breakdowns in an early stage, offers the opportunity to act faster and therefore reduce or even mitigate congestion problems compared to current practice. This paper describes a model based real time short term predictor and its application results, developed to predict traffic states on road segments of typically 50-300 meters on a complete urban and non-urban networks.

This short term predictor is based on four key features; data fusion, real-time estimation of the fundamental diagram, fuzzy traffic state estimation and traffic flow simulation. Traffic flow theory is used to aggregate and fuse data from various data sources (i.e. loop detector data, floating car data and traffic light data) into detailed traffic state estimations per minute. The basis for fusing and completion of data is a macroscopic traffic propagation model within OmniTRANS transport planning software which is also used for near future prediction purposes (up to 10 minutes). As road capacity varies for weather conditions, lightness, number of vehicles et cetera a self-adapting module is used to constantly estimate and update parameters which describe the fundamental diagram for road segments (such as free flow speed, capacity and speed at capacity). Every minute a model-run is performed resulting in actual and near futures traffic states. Subsequently a virtual patrol analyses the measured and modelled data using fuzzy logic to detect incidents on the road network and identify and predict congestion in the near future.

This model-based short term predictor has now been applied with success on the A10 orbital road of Amsterdam (NL) and on a secondary road network of Almere consisting of ten traffic lights. This paper describes the approach used and the results of these two cases.

2 INTRODUCTION

Such as in many European countries, the use of urban and interurban roads is increasing in the Netherlands. As a result road networks are running out of capacity and congestion is increasing as well as the vehicle hours spent. This causes societal disbenefits which increases the urge to solve the problems. One of the solutions is to provide traffic managers insight in the current traffic conditions and near future events to act upon this information and to better inform car users during their trip. The latter can be done with road side units such as Variable Message Signs and within vehicles using in-car equipment. Providing complete current and near future traffic conditions requires advanced methods combining data sources and modelling techniques.

The quality and availability of traffic data has been significantly improved in the last few years. More and more, loop detector data and floating car data becomes available (real time). This offers opportunities for new tools for traffic flow analysis and prediction relevant for operational traffic management services. Being able to detect and predict incidents, e.g. queues, accidents and car breakdowns in an early stage, offers the opportunity to act faster and therefore reduce or even mitigate congestion problems compared to current practice. Supported by the projects CHARM (co-operation between Highways England (UK) and Rijkswaterstaat (NL)) and the iCentrale initiative (Dutch Program in which local, regional and national authorities work together with private parties on better traffic management) a model-based short term predictor has been developed and applied for several real life cases.

This paper describes the theory behind the developed short term predictor and presents the approach and results for two case studies in which the short predictor has been implemented.

3 THEORETICAL BACKGROUND

The development on this model-based short term predictor is primarily based on traffic flow theory within a macroscopic model environment. Additionally, state-of-the-art techniques related to both incident detection and prediction are implemented.

3.1 Fundamental diagram

From Greenshields' observation in the 1930's on, many research has been done on the relation between traffic flow, speed and density of which the mutual relation can be described by a fundamental diagram. The currently considered best approximation of the flow density diagram is described by the inverse-lambda shape. Within this shape the "left-branch" can be seen as the free flow branch and the "right branch" can be seen as the congested branch.

However, although the fundamental diagram describes a theoretical relation between flow and density, "real" traffic does not strictly behave according to this homogeneous behaviour described by the fundamental diagram. Measured data will contain scattered measurements around the fundamental diagram.

Especially the congested branch of the fundamental diagram has shown to be very non-heterogeneous.

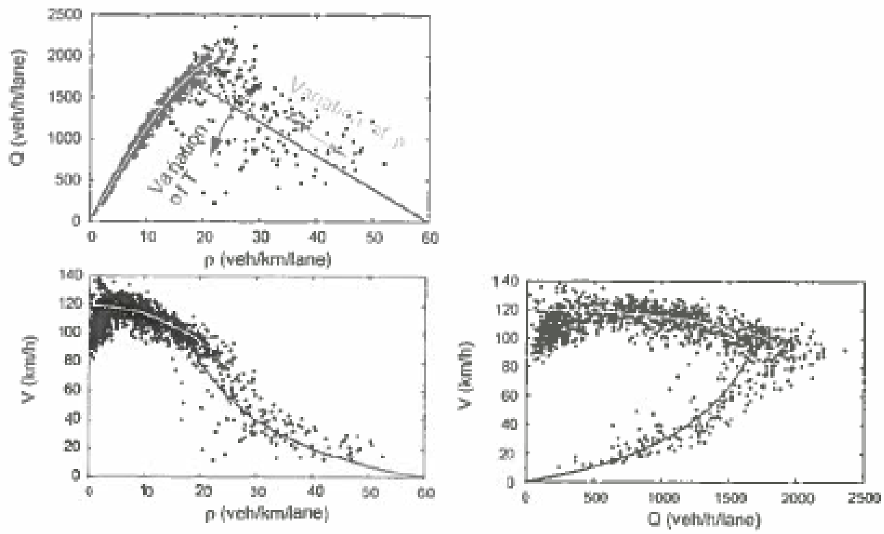

Figure 1: Observations of traffic flow (Treiber and Kesting 2013)

Although the “right branch” of the fundamental diagram is often associated to as the congested branch, Kerner (2003) differentiates into two phases: synchronized flow and the wide moving jam. The wide moving jam is characterized by the upstream movement of the downstream jam front with a constant speed. The downstream front of synchronized flow is normally fixed at a bottleneck.

3.2 Congestion prediction

To determine or predict current and near future traffic states, basically two approaches can be recognized. On the one hand, model-based approaches aim at reproducing traffic situations within a model environment in order to propagate these traffic states for prediction purposes (Kaysi et al. (1993) , Wismans et al. (2014), Vlist et al. (2016). On the other hand, data driven approaches make use of extensive historic data sets to make estimations for near future situation Huisken and Coffa (2000).

Not much research has been published regarding real-time traffic state and congestion prediction. Kaysi et al. (1993) suggested an Advanced Traveller Information System (ATIS) in which historical and updated Origin-Destination matrices were used as input for congestion prediction. It has been suggested by the authors to use 3-Dimensional O-D matrices (with time as a third dimension) to feed a Dynamic Traffic Assignment (DTA) model. This concept has been brought into practice by Wismans et al. (2014) who implemented it for the Assen region in the Netherlands showing the concepts feasibility. However, the authors note that the scalability and the quality of the predictions of this approach are points of attention. Vlist et al. (2016) elaborated further on these points of attention especially regarding the quality of the prediction results. An extensive and continuous calibration process of the network conditions was introduced to improve predictions.

4 SHORT TERM PREDICTOR

Following the approach by Vlist et. Al (2016), the core of the developed short term predictor is a macroscopic dynamic traffic assignment model. Basically, the traffic model is used to process raw traffic data in meaningful estimation and predictions of traffic states for an entire network. Complementary to the model core, supporting submodules have been developed to process raw data, calibrate and improve traffic propagation and allow incident detections.

For this purpose, it has been chosen to elaborate on a model-based approach instead of a data-driven approach as for traffic monitoring and management purposes unexpected events are of special interest compared. A data-driven approach might be strong in predicting expected and average traffic flow patterns and identifying unexpected events in current traffic flow, but such approach is not likely to be successful in predicting the effect of accidents or other unexpected events in advance. Furthermore, the model-based approach has showed to be easily scalable and calculation times are more than fast enough for real time application.

4.1 Framework

First a network must be made available within a macroscopic traffic model. For these developments it has been chosen to make use of the state-of-art dynamic traffic model Omnitrans. Within the traffic model the complete network for which the short term predictor has to operate is included. Subsequently a data processing submodule processes raw data. Data is extracted from its sources, fused with other data sources and mapped on the available network. This processed data is input for the model environment in which demand and supply are calibrated and traffic state predictions are handled.

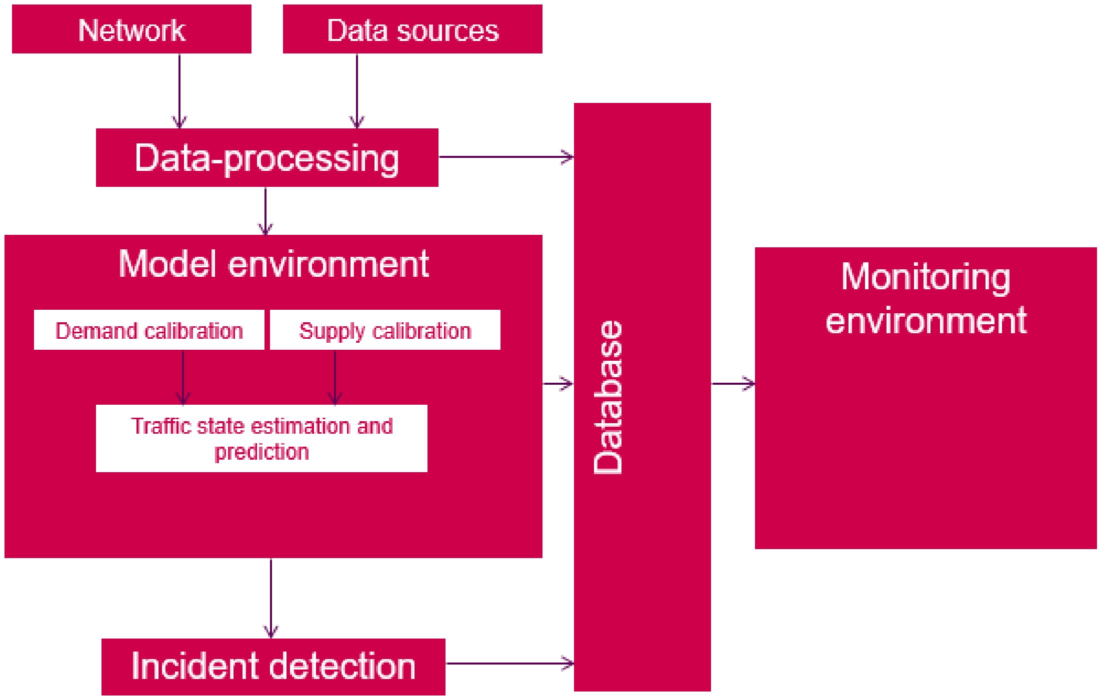

Figure 2: Framework of the model-based short term predictor

4.2 Data processing

The data processing submodule is responsible for extracting and receiving raw real time data from various sources (i.e. loop detector data, floating car data (FCD) or traffic light count data). Raw data is fused and mapped onto the model network, providing minute averaged speed and count measurements for specific segments. The submodule also deals with the differences in latency of the various data sources. Within both use cases in which the short term predictor has been implemented it has been observed that data latency can be up to several minutes. Such latency does have a serious effect on the response time for incident detection.

4.3 Demand calibration

The fused measurement data connected with the model network is used for calibration of the model demand. Flow measurements are used to scale historic origin-destination(OD) matrices in such way that traffic demand fits the demand profiles on predefined locations (typically locations on the borders of the network).

4.4 Supply calibration

As a result of various internal (e.g. speed limits, number of available lanes set by traffic managers or traffic management systems) or external influences (e.g. weather conditions, amount of freight traffic) available supply and traffic behavior differs over space and time. Note that both free flow speed and capacity will decrease in cases of fog or rising freight rates. Due to these dynamics, an assumption of fixed fundamental diagrams for the network would result in inaccurate propagation of traffic over the model network as well as inaccurate detection of incidents and congestion, which affects the quality of traffic state predictions. To avoid the need of various additional data sources like weather conditions and settings of measures as well as the translation of their impact on flow propagation via behavioral models we use measured data to continuously calibrate the fundamental diagrams for each link in the model environment reflecting directly these pre-mentioned influences. Fundamental diagrams are calibrated in three ways depending on the actual traffic conditions: unsaturated free flow, saturated free flow or congestion.

Under unsaturated free flow conditions flows are relatively low and vehicles affects each other driving behavior to a minimum. Such conditions are helpful to determine or update the free flow speed of the particular road segment.

When traffic flow is in (highly) saturated free flow condition, traffic is still in free flow conditions, but individual vehicles affect each other's driving behavior to a large extent. Under such conditions traffic speed has dropped compared to the speed under unsaturated conditions and is around the so-called speed-at-capacity. The speed-at-capacity describes the transition from the "free flow branch" to the "congested branch" within the concave four parameter Van Aerde fundamental diagram (Van Aerde (1995)) used within this approach. Therefore, such conditions are helpful to update the speed at capacity of the fundamental diagram for the particular road segment.

As described in the theoretical background, traffic behaves very non-heterogeneous under congested conditions. This means traffic measurements are widely scattered around the theoretical congested branch as described by the fundamental diagram. Therefore, fundamental diagram calibration is very complex under such conditions. However, if measured data implies congested conditions while previous model predictions did not foresee so, road capacity within the model environment is likely to be overestimated. Under such circumstances capacity can be calibrated in such way that model capacities reflect “real” capacities better. As congested conditions within (in the middle of) a queue are not the result of local lack of capacity, but of a downstream bottleneck, calibration of road capacity is solely done for the downstream road segment of a queue. For urban road networks this approach is also used to calibrate the effect of traffic lights into the model parameters. Capacity of the downstream link of each branch of the intersection is continuously updated in order to include the effect of traffic lights. However, it needs to be remarked that for this purpose it is not aimed for to simulate green light distributions of traffic lights exactly. No real “stop-and-go” effects is simulated around the intersection but more averaged traffic states over time.

4.5 Traffic state estimation and prediction

Within the model environment StreamLine::Madam is used for traffic estimation and prediction. StreamLine::Madam is a macroscopic dynamic traffic assignment model that translates traffic demand on OD-level over time into traffic flows, speeds and densities on a link level for each time-period. StreamLine::Madam reproduces the actual traffic situation (combined with the previously described calibration processes) and calculates traffic states for the short term prediction horizon which is typically 1 to 10 minutes. The streamline::Madam submodule consist of four steps which are illustrated in Figure 3.

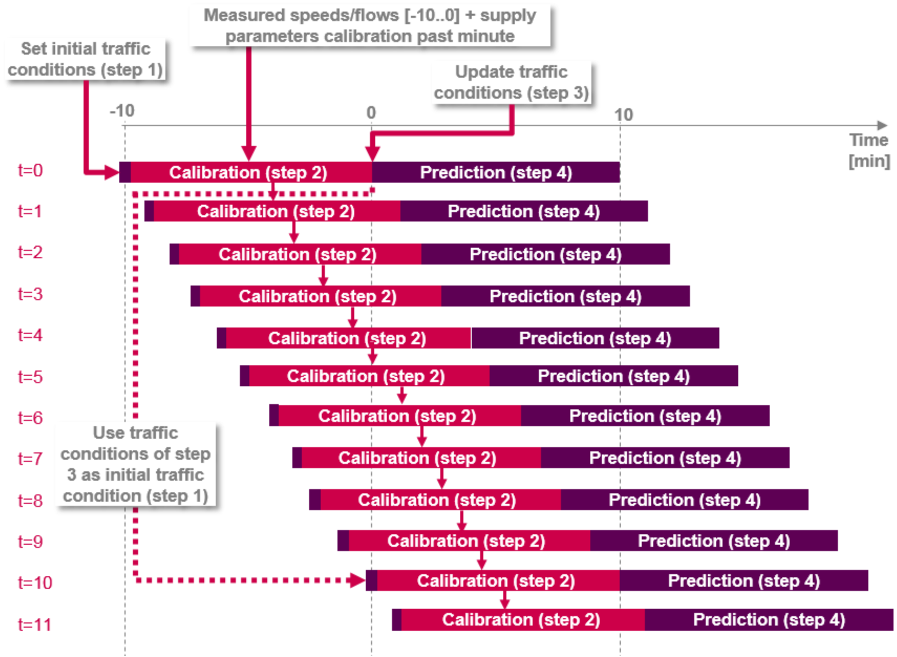

Figure 3: The process of traffic state estimation and prediction

The first step is to set initial traffic conditions so that traffic states for the full networks matches traffic states based on measurements. Compared to the traditional network loading process of dynamic models an alternative approach is used for this step to save valuable calculation time. With a so-called warm start the product of step 3 of a previous simulation run is placed directly on the network as starting point of the simulation. After this process StreamLine::Madam propagates traffic for a ten minute calibration process in step 2. Note that the calibration step is done using previous minutes for which measurements are known. After each minute of propagation the supply calibration submodule as described before updates supply parameters. Consecutively, after all supply parameters have been optimally updated, in step 3 traffic states on the network are updated using most recent measurements. At last, StreamLine::Madam propagates traffic for second time for typically a ten minute prediction horizon.

4.6 Incident detection

Besides a traffic state prediction the short term predictor does also assess predictions and measurement data in order to detect incidents on the network. This is primarily done by comparing live measurement data with previously predicted traffic states. For this comparison first a classification of the data is made. Each road segment is classified with a likelihood of being congested. If traffic flow is clearly within the congested branch of the fundamental diagram this likelihood is set to 1. On the other hand, if traffic flow is clearly uncongested the likelihood is set to 0. As traffic behaves non-heterogeneous around capacity fuzzy rules including speed and flow are needed to classify the likelihood of road segments on which traffic is around capacity.

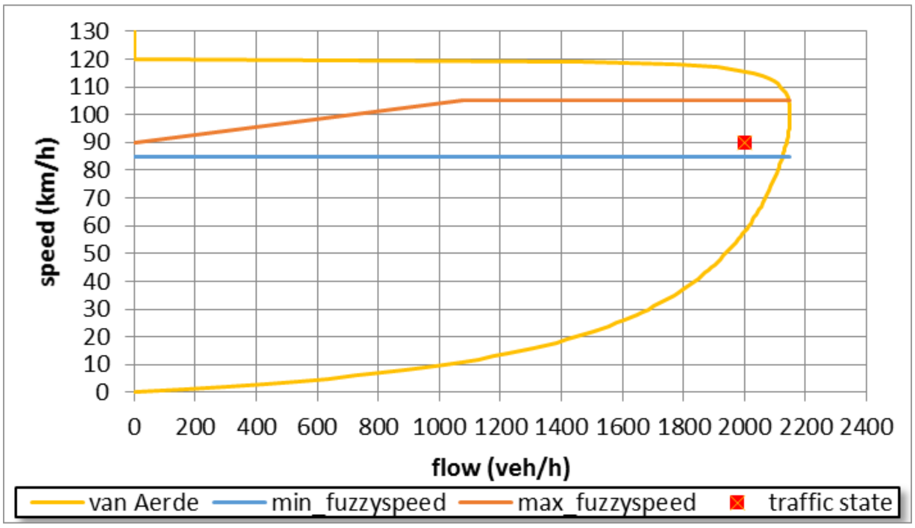

Figure 4: The likelihood framework used for incident detection

Using this qualification incident detection is performed for any road segment for which measured data is available. For each road segment the measured likelihood is compared to the predicted likelihood. The predicted likelihood is an aggregation of all previously predicted likelihoods of previous model runs in which old predictions are weighted less then more recent predictions. In cases that the measured likelihood of congestion does not fairly match the predicted likelihood, a probability is calculated of an incident has happened on this location. For example: If predictions did only show free flow traffic on a particular road segment while recent measurements definitely show a congested situation, the road segment is flagged with a high probability that an incident occurred on that location.

4.7 Online monitoring environment

Both traffic state predictions and incident detections are communicated using an online monitoring environment. In this environment speed and flows are visualized for current and near future time steps for the full network. Figure 5 shows a visualizations of this monitoring environment for both use cases in which the short term predictor is implemented. Complementary to this online monitoring environment traffic state predictions and incident detections can be communicated to traffic management centers in which they can be helpful in active traffic measurements and within decision making processes related to scenario planning.

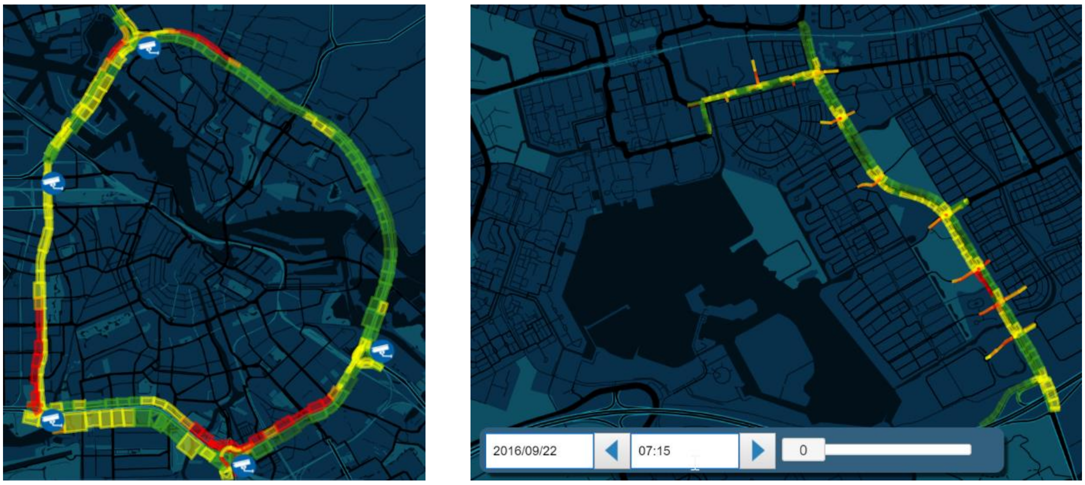

Figure 5: Examples of the online demonstration tool for traffic predictions

5 CASE STUDY

The short term predictor has been implemented in two use cases. These use cases come forward from two projects in which this short term predictor has been developed and implemented. Supported by the CHARM-project (co-operation between Highways England (UK) and Rijkswaterstaat (NL)) a highway use case has been set up. Besides, within the iCentrale initiative (Dutch Program in which local, regional and national authorities work together with private parties to improve traffic management services) the short term predictor has been implemented within an urban road environment. Both use cases differ on scaling and complexity. Where on highways road capacities are the major cause for delays and congestion, congestion on urban road networks is primarily caused by the impact of (signalized) intersections. With the presence of intersections, lower data availability as well as data quality (e.g. as a result of penetration rates of floating car data and smaller absolute numbers of vehicles), larger routing options, the presence of actual origins and destinations (e.g. vehicles parking or departing on road segments) and mixture with various types of road users on urban road networks, such environments are far more complex than highway environments.

The highway use case includes a network of highways around the city of Amsterdam in the Netherlands: the A10 orbital road. For functional reasons the network includes not only the highway sections itself but complementary it includes all on- and off ramps and connections to connecting highways. The highway use case has been implemented in a live environment. Both streaming loop detector data as floating car data has been processed in real-time and traffic has been monitored in a live environment. Figure 6, visualizes the selected network.

Figure 6: Use case network (left: the Amsterdam A10 orbital road highway use case, right: the Almere urban road corridor use case

The urban road use case consists of a corridor including multiple intersections of which 10 are signalized. The network describes a major corridor from the A6 highway towards the city centre of the Dutch town Almere. In contrary to the highway use case, the urban road use case has been implemented in an offline environment. No live data is processed within this use case. However, for simulation purposes a live environment is imitated in which no live but historic data is processed resulting in streaming loop detector data, floating car data and traffic light data.

6 RESULTS

In both use cases similar evaluation indicators have been used to assess the quality of the output of the short term predictor. These indicators combined give a good qualification of the module performance on its predictive ability.

6.1 Evaluation framework

The evaluation framework consists of two indicators. A global network indicator based on the number of kilometers of congestion in the complete network and a network indicator assessing the accuracy in predicting the correct locations of congestion as well as non-congestion. The evaluation which is done afterwards focusses on the extent in which the predictions match the measurements. Complementary to these assessment indicators the quality of the incident detections is assessed, focusing on the current traffic conditions.

The assessment of accuracy of the model is calculated by comparing the predicted states for the single road segments with the measurements. Road segments of typically 50-300 meter are used for the calculation of the accuracy. Therefore, similar to the global network indicator the results of the short term predictor are ex post compared to measurements. However, in contrary to the global network indicator the statistical framework helps to assess whether or not congestion is prediction on the correct locations on the network. Assuming a single road segment can either be congested or non-congested, the comparison of the results of the short term predictor with measurement can result in four possible combinations. A congested prediction can either be true or false and so can a non-congested prediction be, resulting in four quadrants:

Q1: True Positive
Q2: False Positive
Q3: False Negative
Q4: True Negative

	Measured congestion	**No measured congestion**
Modelled congestion	*Q1 (# segments)*	*Q2 (# segments)*
No modelled congestion	*Q3 (# segments)*	*Q4 (# segments)*

For every link segment within the network the indicator determines for each modelled minute to what quadrant it belongs. From these segment results an overall evaluation of the accuracy of the model can be calculated:

Accuracy = (Q1+Q4)/(Q1+Q2+Q3+Q4)

6.2 Highway use case

The short term predictor has proven to be successful in reproducing and predicting network traffic states. Where congestion is varying over time and space during the day, the short term predictor is able to follow and predict these patterns. Figure 7 visualizes the +6 minutes predictions of the short term predictor against the measured data in terms of congestion kilometres per minute. As it can be seen, the module is well able to predict fluctuation in congestion over the day.

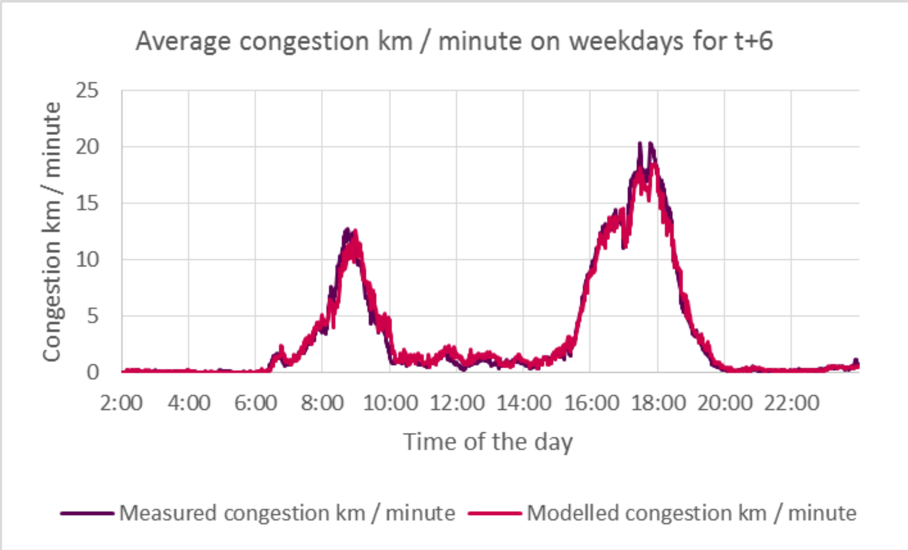

Figure 7: Measured versus predicted (+6 minutes) congestion kilometres on the modelled network

Besides a successful reproduction of traffic patterns, the short term predictor has also proven to do so with a relatively high accuracy. The accuracy of the model has been determined for each morning peak for a one-week period for both three and six minute predictions. These evaluation results are presented in figure 8. Overall, it can be seen that the accuracy of the module is reasonably high with an average accuracy of over 90% for both three and six minute predictions. For morning peaks an average accuracy is reached of 95% for three minutes prediction and 93% for six minute predictions. For evening peak periods the accuracy level is only slightly lower with an average accuracy of 92% for three minute and 91% for six minute predictions.

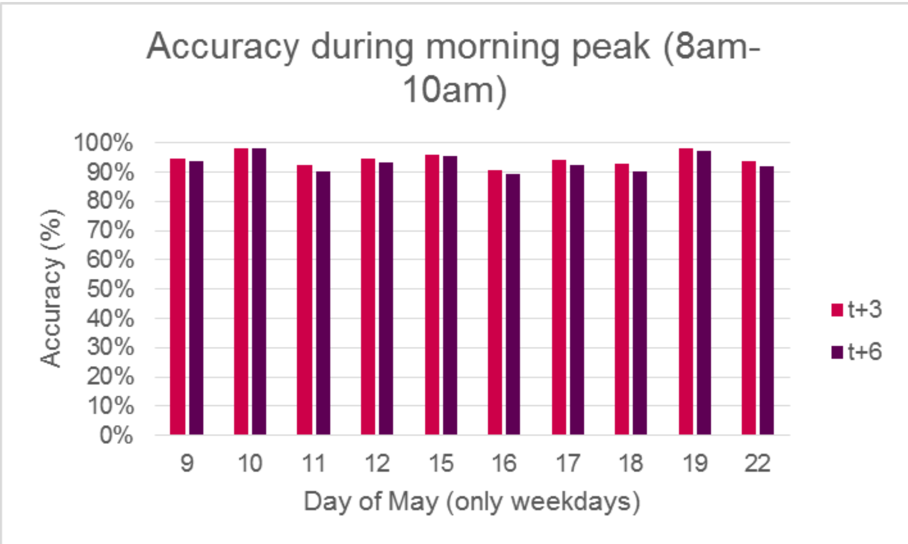

Figure 8: Accuracy figures for morning peaks on weekdays

For one of these morning peaks a more detailed analysis of the accuracy is performed. In figure 9 the four quadrants used for the accuracy calculation are visualized for a single morning peak. The sum of the true positive and true negative is the accuracy level as presented before. From figure 9 can be concluded that the network was largely in free flow condition during this particular morning peak. Furthermore, the number of false positives and false negatives are reasonably equally distributed over time, which means that the accuracy does not differ depending on the level of congestion within the network. However the congestion was not always predicted on the exact correct position, although the total number of congestion kilometers shows an accurate match. Further analysis shows that in some cases the model predicts congestion to occur a few segments downstream of its actual location and in some cases the model slightly under- or overestimates the shockwave speed of a queue.

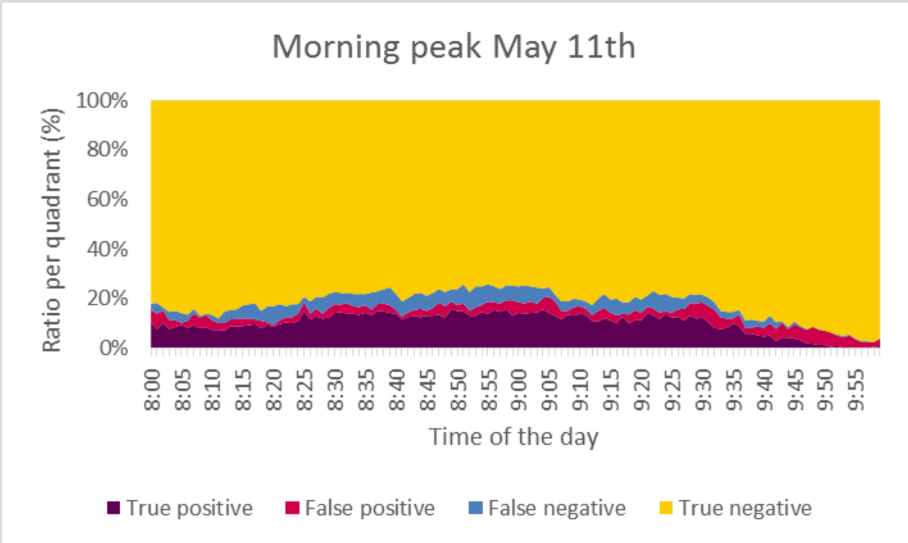

Figure 9: Detailed visualization of accuracy figures for 11th may morning peak

6.3 Urban road use case

Just as for the highway use case the short term predictor has proven to be able to reproduce actual traffic situations in the urban case reasonably well. The accuracy on an average morning peak has been calculated to be 96% for the current situation. This fairly good reproduction of the actual traffic situation offers an excellent starting point for the prediction horizon. Here it is seen, that accuracy decreases the longer term the prediction is. For 10 minute predictions the accuracy level is still 88%. The decrease in accuracy is best to be explained as a result of uncertainties in intersection delay. Intersection capacity is to be calibrated by the supply calibration, but due to varying green light distributions in the prediction horizon it might differ more for coming minutes than is the case for capacities calibrated for highway segments. Furthermore, it can be expected that the variation in local demand in urban networks differs more as a result of the influence of intersection on the propagation of traffic, larger routing options and smaller absolute numbers of cars on specific segments resulting in larger relative deviations, which earlier also addressed as factors increasing the complexity on urban networks.

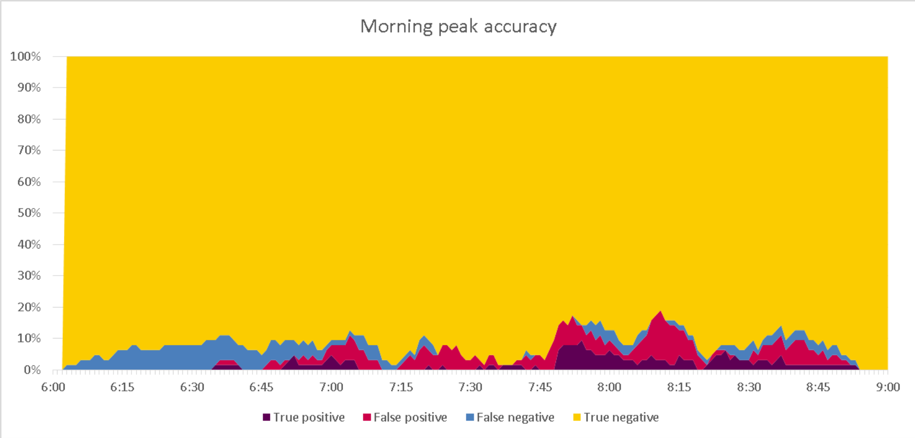

Figure 10: Detailed visualization of accuracy figures for an average morning peak

7 DISCUSSION AND CONCLUSIONS

Implementation of our model based short term predictor in two use cases has proven our developments to be successful in traffic state estimation and prediction. From various data sources the short term predictor successfully provides a complete and consistent picture over space and time for a complete network offering the opportunity to serve as the common operational picture for traffic management purposes. The model-based approach handles regular but also irregular events affecting the supply parameters of the network and the network demand. Furthermore, it allows the possibility to not only detect and predict near future traffic states based on actual situation, but as well calculate the effects of multiple what-if scenarios (i.e. traffic management scenarios).

With an accuracy of around 90% (and higher) the short term predictor has now been implemented in both a highway as an urban environment. To improve performance of the module significantly one should focus on two domains. On the one hand, experience within these two use cases has shown, that although supply calibration has been given major attention, the quality of traffic demand is a crucial element that can affect prediction quality. More advanced demand calibration algorithms have however showed to require too much computational effort within a live performing algorithm. Therefore, further developments on this short term predictor will definitely take this issue into account.

A second domain which needs further attention, is data availability and latency. For both use cases calculation of traffic state estimation and prediction was very fast and convenient for live applications. However, latency on traffic measurement data has showed to be a serious aspect. Depending on the data source, latency of up to several minutes have been observed. As such measurement data forms the basis for any approach (model-based or data-driven) for traffic state estimation or incident detection, such high latencies significantly affect the response time within incidents are noticed and measures can be taken. Therefore it is advised that more effort is committed towards making data available faster.

BIBLIOGRAPHY

Huisken, G. and Coffa, A. (2000). Short-Term Congestion Prediction: comparing time series with neural networks, IEE Conference Publication 472, pp. 60 – 63.

Kaysi, I. Ben-Akiva, M.E., and Koutsopoulos, H.N. (1993) Integrated approach to vehicle routing and congestion prediction for real-time driver guidance. Transportation Research Record, 1408.

Kerner, B. S. (2003). "Three-phase traffic theory and highway capacity." Physica A 333: 379-440.

Treiber, M. and A. Kesting (2013). Traffic Flow Dynamics, Springer.

Van Aerde, M., 1995, “Single regime speed-flow-density relationship for congested and uncongested highways”, 74th Transportation Research Board Annual Conference, Washington, D.C., Paper No. 95080

Vlist, M. van der, L.J.J. Wismans, P. van Beek & L.C.W. Suijs (2016). Virtual patrolling. *Transportation Research Procedia*, 14 (2016), 3370-3379. doi:10.1016/j.trpro.2016.05.289

Wismans, L.J.J., E. de Romph, K. Friso, & J. Zantema (2014). Real-time traffic models, decisions support for traffic management. Procedia Environmental Sciences, 22, pp. 220-235, doi:10.1016/j.proenv.2014.11.022