Incident Detection and Incident-Impacted Traffic Prediction for Urban Road Networks, Application to GrandLyon

Laura Wynter, Sebastien Blandin
IBM Research Singapore Collaboratory, Singapore
lwynter@sg.ibm.com

Barry M Trager, Yichong Yu
IBM Research, Yorktown Heights, NY USA

Yiannis Kararianakis
Arizona State University, Tempe, Arizona, USA

Jean Coldefy, Dimitri Marquois
ITS and Mobility services unit, GrandLyon, Lyon, France

Thomas Baudel
IBM France Lab & Efficacity Institute, France

ABSTRACT

This paper describes joint work done by IBM Research (development of the solution) and GrandLyon (assessment of the solution) for real-time incident detection and traffic prediction on an urban road network in the presence of incidents. The detection and prediction methods are integrated and tested in a pilot study on live real-time traffic data in Lyon Metropolis, France. Numerical results and analysis are provided. The benefits of faster and more accurate incident detection include faster incident clearing times and more accurate and timely information to the public so as to avoid affected areas on the road network.

1. INTRODUCTION

Real-time traffic prediction is a critical component of modern road traffic management systems. Indeed, for most traffic management functions to perform adequately, the raw real-time traffic data is obsolete by the time it is received. Hence, real-time traffic prediction, which forecasts the future traffic state from a few minutes to one hour into the future, using real-time data, is needed for intelligent transportation systems applications, such as real-time route guidance, adaptive traffic control and advanced traveler information systems.

While traffic prediction technology is highly effective in most network conditions, when an incident occurs on the road network, the statistical traffic prediction methods employed by spatiotemporal time series prediction methods such as those in Min and Wynter (2011) and Kamarianakis, Shen and Wynter (2012) may not perform adequately, especially at the beginning of the incident. Indeed, parameters used by time series methods are calibrated on historical conditions, which include only a limited number of incidents. However, the number of incidents occurring on the road network at any particular place and time is relatively low. We therefore propose a different approach comprised of an incident detection method, IDM, to automatically detect incidents from the raw traffic data, and an incident-specific traffic prediction method, called Incident Duration Prediction, or IDP, for the context of road traffic prediction in the presence of incidents.

This paper presents the methods developed by IBM Research and calibrated and tested for the incident detection and duration prediction components during the course of the Optimod'Lyon project (http://www.optimodlyon.com/en/). The Optimod'Lyon project consisted of a consortium that tested, among other tasks, multiple solutions from several companies for various forms of traffic prediction. This paper presents the assessment by GrandLyon of the methodology developed by IBM Research on the Lyon Metropolis road network. The benefits of faster and more accurate incident detection include faster incident clearing times and more accurate and timely information to the public so as to avoid affected areas on the road network.

In the next section, we discuss the methodology tested for incident detection. In section 3 we present the methodology developed for traffic prediction in the presence of an incident, and in section 4 we present results obtained for the Lyon road network. We conclude in section 5 with future work.

2. INCIDENT DETECTION METHODOLOGY

Incident detection algorithms based on traffic data from fixed sensors can be classified into three main types, (see, e.g. Parkany and Xie, 2005).

- comparative algorithms based on decision-trees that use a set of decision variables and a set of (possibly location specific) thresholds to classify a traffic state in a particular location as incident free, tentative incident or incident;
- time-series approaches, based on accurate forecasting models of traffic variables; in this case incident detection occurs when detector measurements deviate significantly from the corresponding forecasted values;
- artificial intelligence algorithms typically based on fuzzy logic and/or neural networks.

The proposed method, IDM, focuses on algorithms of the first type. One of the most successful examples is the California algorithm (Payne and Tignor, 1978); its structure is depicted in Figure 1.

In our proposed method, DOCC denotes the downstream occupancy observed at time t (i.e. occ_t^d), OCCDF represents spatial difference in occupancies between a set of (upstream and downstream) detectors (i.e. $occ_t^u - occ_t^d$), where the superscript u indicates an upstream detector, and OCCRDF is the relative spatial difference in occupancies (i.e. $\left(occ_t^u - occ_t^d\right)/occ_t^u$). In this case, downstream and upstream refer to a reference location, such as the potential incident location. T1, T2 and T3 represent thresholds for DOCC, OCCDF and OCCRDF, respectively, and the state variable takes on four values: 0 (incident-free), 1 (tentative incident), 2 (incident occurred) and 3 (incident continuing).

One expects that effective calibration of the above algorithm essentially requires substantial spatio-temporal variability in T1, T2 and T3: for instance T2 is expected to differ substantially when the downstream detector is placed in a bottleneck as opposed to the case that the bottleneck is located upstream. The tedious calibration procedure required for the effective implementation of algorithms as the one presented above has been reported in a number of reports, including Martin et al. (2001).

Our method aims to overcome this issue by requiring calibration for a small number of parameters while accounting for the spatiotemporal variability of the variables that are included in the decision tree.

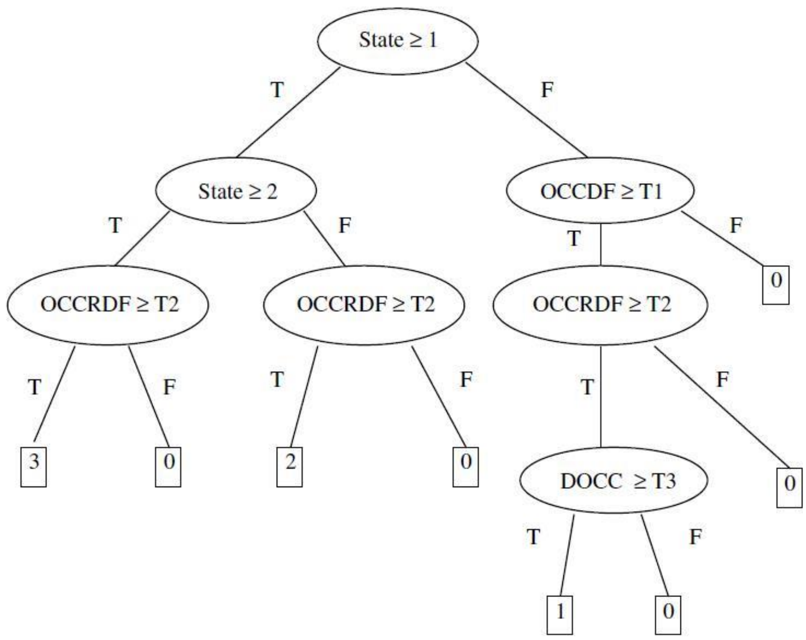

Figure 1. Decision tree for California algorithm 7 (Payne and Tignor, 1978).

In what follows we illustrate our methodology using a decision tree algorithm that is based on data from a single detector. The algorithm uses two decision variables and is depicted in Figure 2. In this case the variable *state* takes on 7 values: 0 (incident-free), 1 (tentative incident downstream), 2 (incident occurred downstream), 3 (incident continuing downstream), -1 (tentative incident upstream), -2 (incident occurred upstream) and -3 (incident continuing upstream). Essentially the algorithm seeks sufficiently large shocks in occupancies which bring occupancy levels outside a band to trigger detection. Occupancy is used in this example as it leads to superior detection performance compared to other decision variables; however, the proposed method is not dependent on this choice and other variables (such as the ratio of volumes to occupancies or speed) can be used as well.

For each measurement location $diffocc = occ_t - occ_{t-1}$ represents shocks in occupancy observed at time *t* whereas $occ = occ_t$ stands for observed occupancy levels. In contrast to the algorithm presented in the previous section, the thresholds are time-varying and defined universally for all measurement locations of the network, based on six parameters: τ_1, τ_2 with $\tau_1 < \tau_2$ are chosen quantiles for shocks in occupancies (Figure 3), τ_3, τ_4 with $\tau_3 < \tau_4$ are chosen quantiles for occupancies and the last two parameters, denoted λ_1, λ_2, control the degree of smoothness of the (time varying) functions, Q_1, Q_2 and $Q_1(\tau|t)$, is the conditional quantile function for shocks in occupancies whereas $Q_2(\tau \mid t)$ is the conditional quantile function for levels in occupancies. These functions can be constructed using the nonparametric quantile regression framework presented in Koenker (2005, Chapter 7). The width of the bands that are based on Q_1 (Q_2), depend on the difference $\tau_2 - \tau_1$ $(\tau_4 - \tau_3)$ and the variability of the location-specific decision variables; hence are time-dependent and location-specific. Figure 3 depicts an example of such decision bands, based on data from a measurement location; it can be observed that the width of the band is substantially reduced during the early morning as occupancies display less variability during this period. Similarly, the corresponding band for a measurement location with substantially more variable traffic dynamics than the one shown in Figure 3, would have been wider. It is worth stressing that a simplified approach of band construction would be based on fixed multiples of (time dependent) standard deviations; such a choice is not expected to lead to satisfactory results when the decision variables are asymmetric/skewed.

For a decision-tree algorithm that is based on *k* decision variables the presented methodology requires calibration of at most *2k+2* parameters as opposed to location-specific parameters. In general, such calibration can be performed along the lines of Staphanedes and Chassiakos (1993), based on a set that contains both incident and incident free traffic data: given a maximum allowed false alarm rate, the chosen parameters should maximize detection rate. A grid-search procedure can be invoked for such purpose. In the example displayed above, τ_3 and τ_4 can be chosen to comply with a set of reported incident durations (as incident duration is dictated by occupancy levels that lie outside their band after a significant shock has been detected); the lambdas can be chosen using a roughness penalty approach, as in Koenker (2005) and τ_1, τ_2 so that the detection rate is maximized for a given false alarm rate.

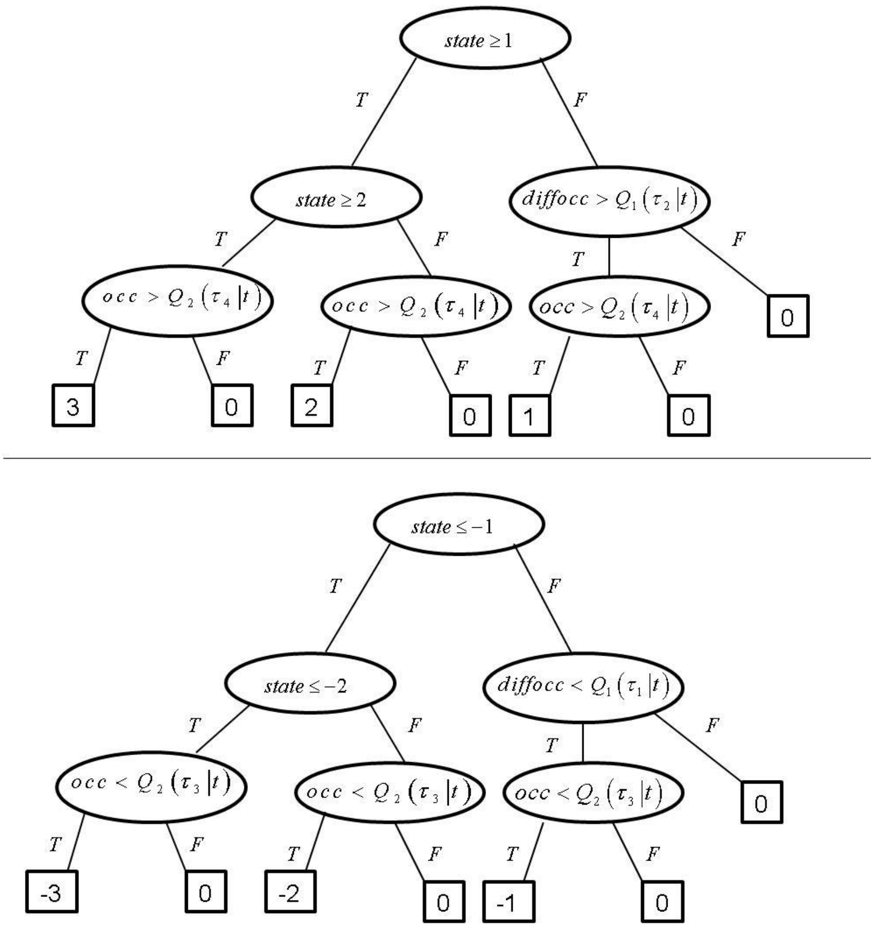

Figure 2. Decision trees for a customized algorithm that is based on two decision variables from a single detector.

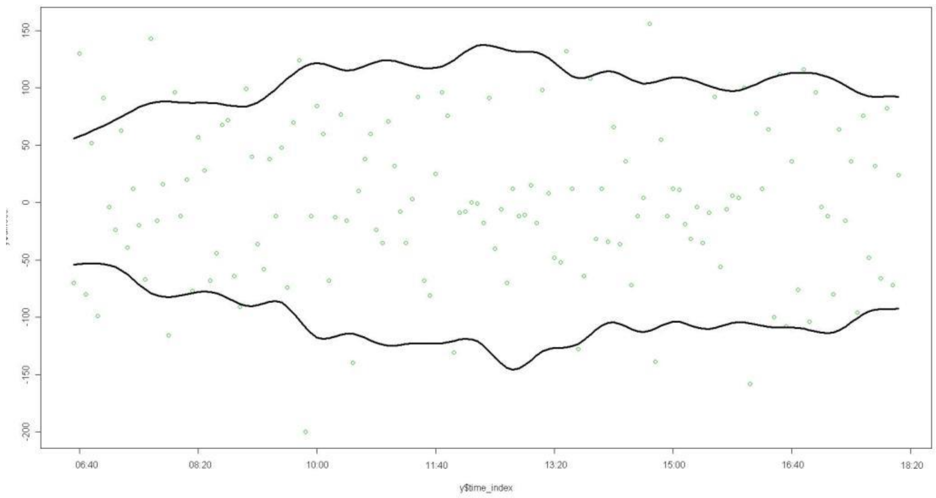

Figure 3. Decision-bands for shocks in occupancies based on nonparametric quantile regression for a measurement location in the road network. The depicted traffic data correspond to a 12-hour period that contains morning peak.

3. INCIDENT DURATION PREDICTION METHODOLOGY

IDP is a parsimonious traffic prediction method focused on predicting traffic conditions resulting from an incident, at the location of neighboring detectors. IDP can predict occupancy, flow, or speed levels. Additionally, metrics of interests for traffic control operators can be computed in real time, e.g. the predicted incident duration, defined as the length of the time period at the end of which traffic conditions return to normal conditions on neighboring detectors.

IDP is triggered by the observation of an incident by a traffic operator, or by the automated detection of an incident. Data types required by IDP include

- Static road features at the location of the incident; road category, number of lanes, speed limit, etc.
- Traffic data at the location of the incident; occupancy (can be changed to speed or volume if needed).
- Incident features; start time of the incident, type of accident, number of vehicles, number of lanes closed. It is clear that some of these features would not be available online when the incident is detected by IDM.

IDP produces a time series of traffic conditions at the location of the incident, expressed as an occupancy value for every 6 min over the next hour. This prediction is updated as new data becomes available, i.e. each 6 min.

We consider a piecewise affine (PWA) time-series model of local incident impact in which traffic occupancy at the location of the incidents goes through 3 different phases before returning to normal traffic conditions;

- Upward phase during which occupancy increases from normal traffic conditions following the occurrence of the incident.

- Flat phase during which occupancy stays stationary around a large value increasing with the local impact of the incident.

- Downward phase during which occupancy decreases toward normal traffic conditions as the incident clears around this location.

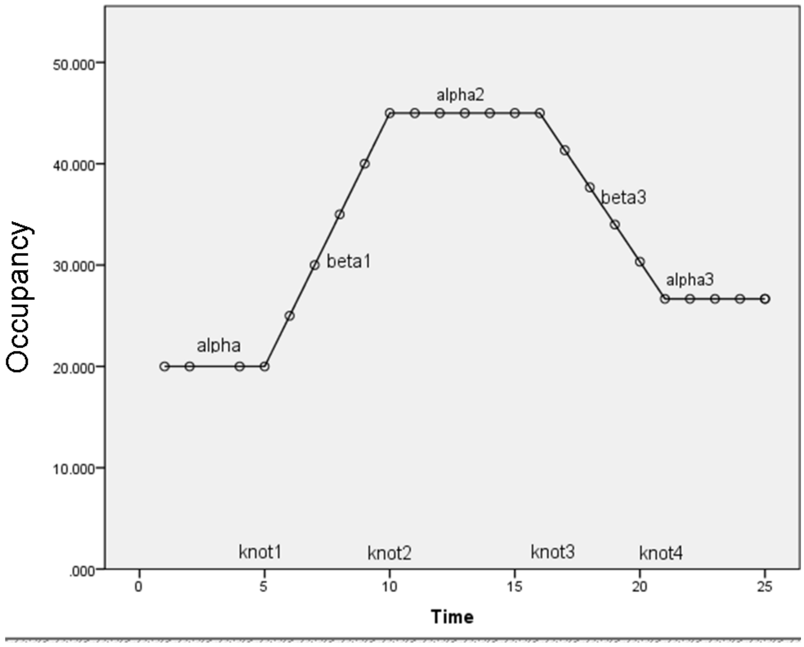

Figure 4. Piecewise affine (PWA) time-series prior of local incident impact.

From this incident model, a number of metrics can be extracted regarding a given incident, illustrated in Figure 4:

- The difference between knot 4 and knot 1 indicates the duration of the incident.

- The difference between alpha2 and alpha indicates the amplitude of the incident.

- The product of these two quantities characterizes the local impact of the incident.

IDP is composed of three components, a nonlinear regression model producing a piecewise affine prior of the local incident impact, a nonlinear classification model assessing the impact of specific types of incidents, and an online update model integrating streaming data.

The first IDP component is focused on the calibration of the PWA model described in previous section. Based on historical occupancy data observed during incidents, this component runs a number of constrained nonlinear models as well as unconstrained nonlinear model with various initial conditions in order to produce optimal values for the parameters alpha, alpha2, alpha3, beta1, beta3, as well as knot1, knot2, knot3, knot4 for each historical incident.

The second component of IDP is focused on producing a value for the set of parameters of the PWA for each type of incident, with the intent that when an incident is observed, a prediction can be made based on this set of parameters. Only minimal incident features available at start time are used, i.e. number of lanes, speed limit, start time, and occupancy value. Based on this history of calibrated events, a neural network model produces the optimal PWA parameters for each type of incident. This produces the first prediction from IDP which is made the start of the incident.

As the incident progresses, new data becomes available. The third component of IDP is focused on the posterior update of the initial prior of IDP, every 6mn as new data becomes available. A Bayesian update scheme is proposed which numerically computes a maximum likelihood estimate of the PWA parameters at each time step. The numerical precision of the Bayesian method can be tuned according to computational requirements.

The method is illustrated in the figure below, the green curve denotes the prior, the blue curve denotes the past data, the grey curve denotes the future data, the orange curve denotes the prediction made at the current time.

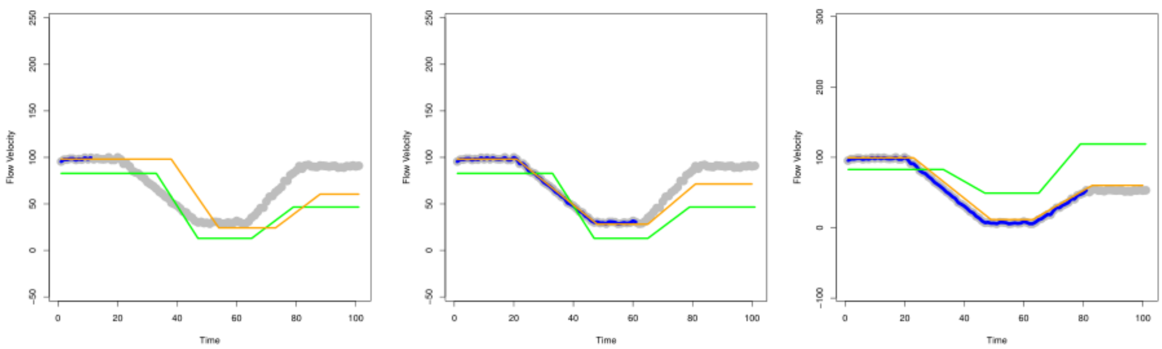

Figure 5. Bayesian update of local incident impact prior.

4. RESULTS

The following three figures illustrate the IDM incident detection method on occupancy data from Lyon.

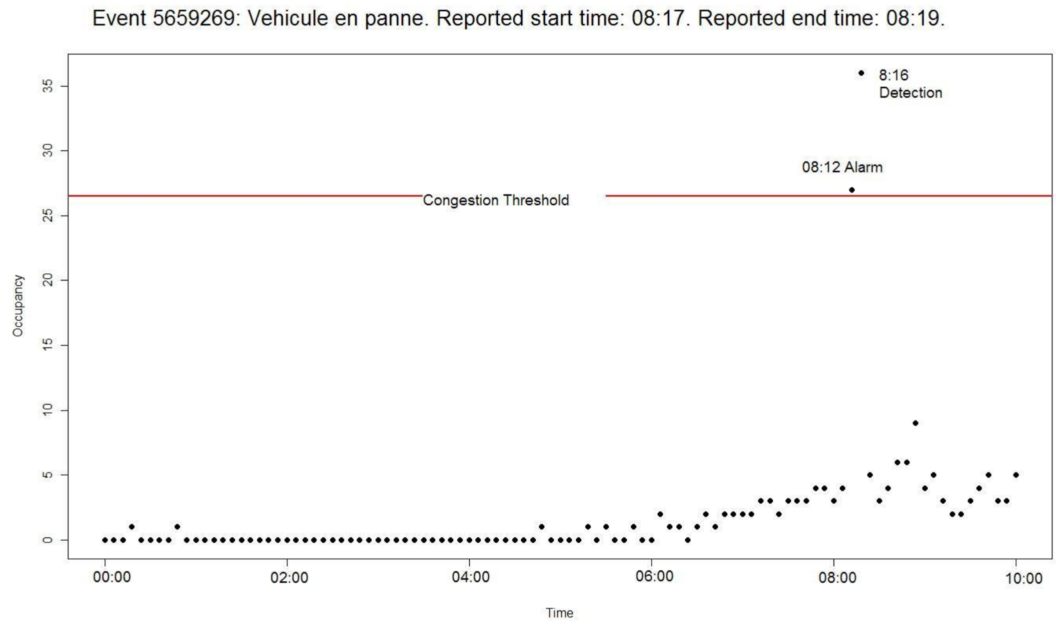

Figure 6. Incident detected by IDM before reported by operators. Higher frequency traffic data would enable even earlier automatic detection.

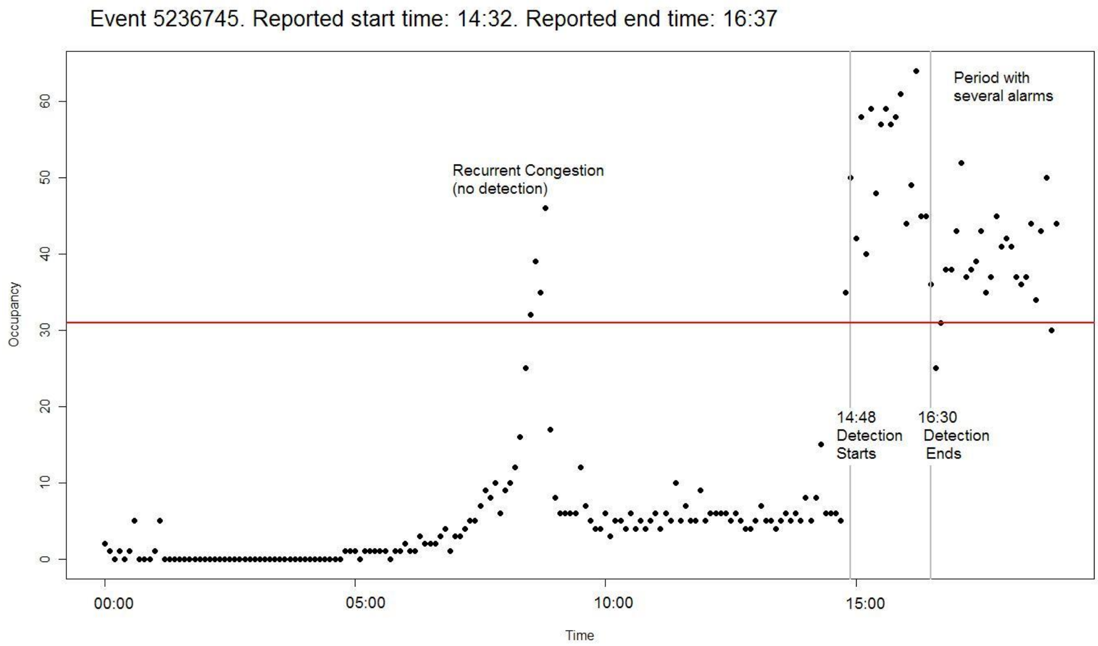

Figure 7. Incident detected by IDM when traffic impact occurs, after the official start time, which was not used in the detection.

Next, we illustrate accuracy results of the IDP component on data from Lyon, France, consisting of 627 incidents. 4/5 of the data is used for training and 1/5 of the data is used for validation.

We compare below IDP (blue curve) against the following models:

- Decision tree trained using IDM duration of the incident as a response. (black curve)

- Decision tree trained using fitted duration knot4-knot1 from PWA model described in previous section as a model. (red curve)
- Multivariate decision tree using all fitted parameters from PWA model as a response. (green curve)
- Proposed method using a multi-layer neural network.

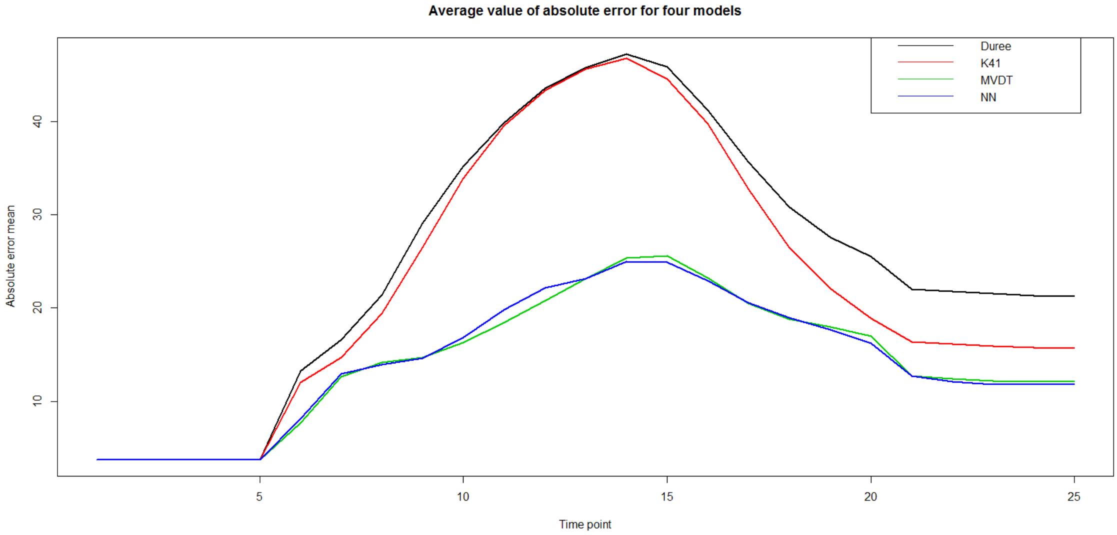

Figure 8. Comparison of average error of IDP across 4 methods.

Note that the multivariate decision tree approach provides results of comparable accuracy to the proposed neural network approach in terms of mean error. The neural network approach tends to have a lower error variance and hence was preferred to the multivariate decision tree.

Finally, we present the relative performance of IDP and the advanced spatiotemporal method, TPT, from Kamarianakis, Shen and Wynter (2012) for incidents, for a 1 hour horizon comprised of 10 6-minute forecast points, for incidents reported by IDM. Figure 9 illustrates the mean absolute error (MAE) of the two methods applied to the same traffic detector over all incidents that were detected during one month, when IDP is run one time period after incident detection. Time point TP 10 indicates the detection time of the incident, and hence Figure 9 assumes that IDP and TPT are run at the next time point after detection, i.e. 6-minutes from the detection time, i.e. TP 11. The following two figures, 10 and 11, illustrate the mean absolute error over the same month when IDP is run two and three (resp.) time periods after incident detection. Note that the benefits of IDP are greatest at the start of the incident, after which point the spatiotemporal-based prediction method of TPT is able to obtain a higher accuracy.

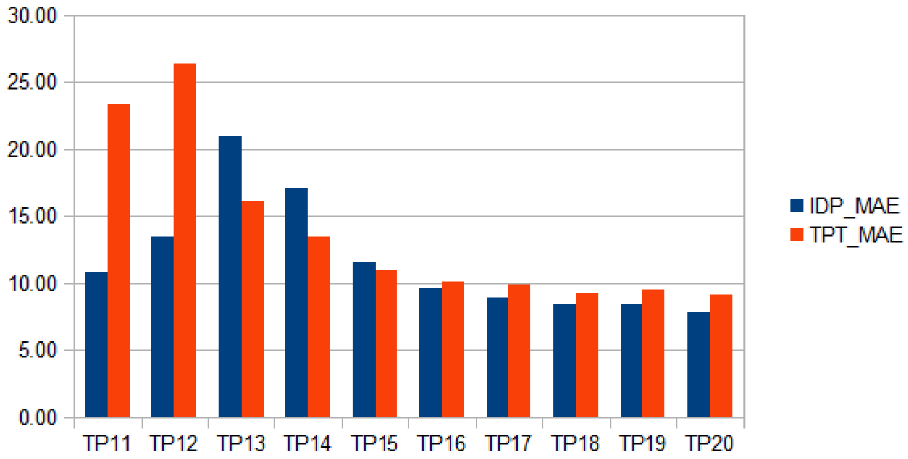

Figure 9. Comparison of mean absolute error of IDP versus TPT over 10 forecast time periods (1 hour), i.e. IDP is run one time period after detection (TP11). IDP has significantly lower error for first two prediction time horizons.

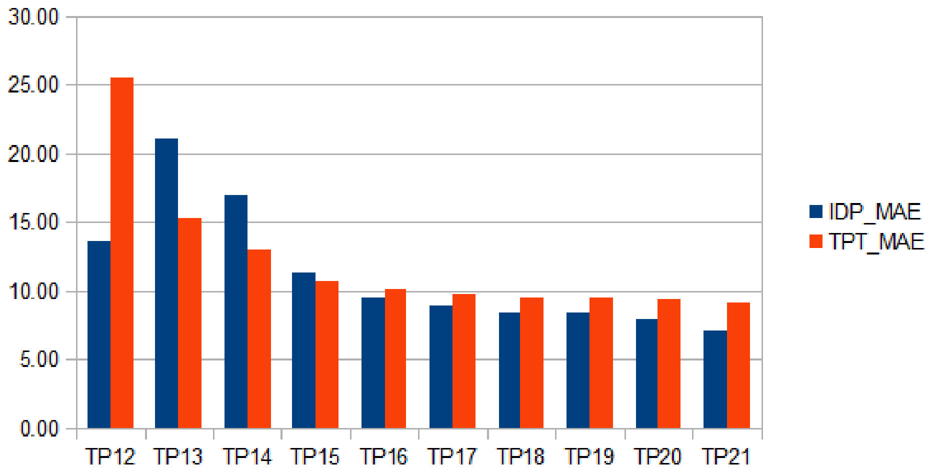

Figure 10. Comparison of mean absolute error of IDP versus TPT over 10 forecast time periods (1 hour), i.e. IDP is run two time period after detection (TP 12).

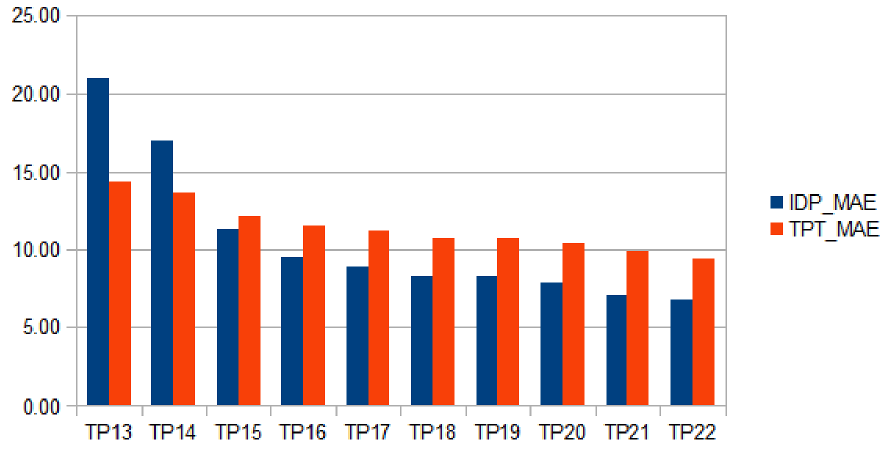

Figure 11. Comparison of mean absolute error of IDP versus TPT over 10 forecast time periods (1 hour), i.e. IDP is run three time period after detection (TP 13). By this time, TPT using spatiotemporal prediction is able to offer a higher accuracy on average.

To further illustrate the difference between IDP and TPT across a large set of IDM-detected incidents, we present below the histogram of point-wise errors across all prediction times, when the prediction is made at the time at which the incident is detected. We note that the histogram of errors is centered below zero indicating that the majority of the time the IDP obtains a lower error than the spatiotemporal approach of TPT in the presence of an incident, across all prediction times.

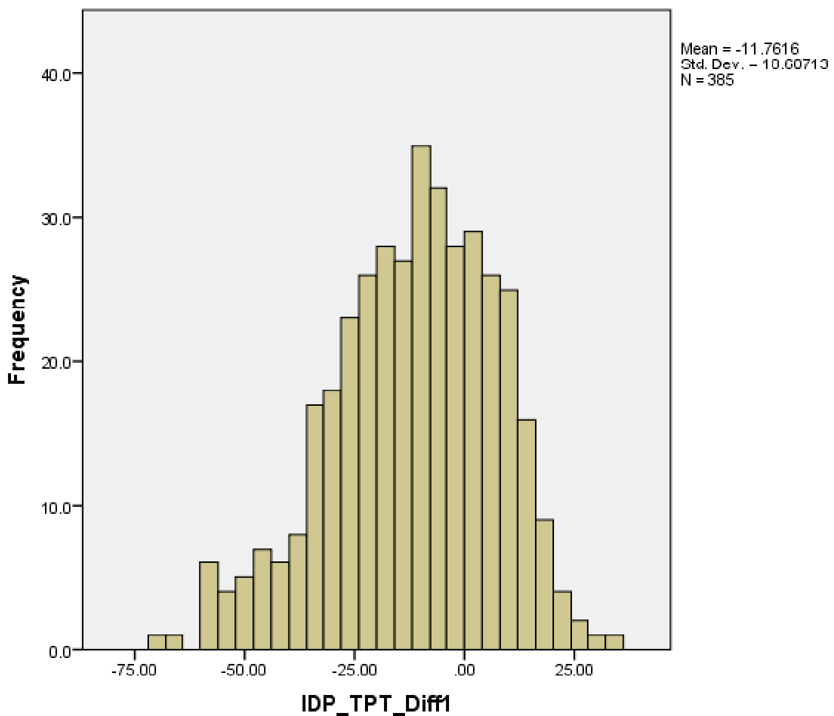

Figure 12. Histogram of error of IDP – error of TPT over all 10 forecast time horizons. Negative values indicate that IDP has lower error.

5. CONCLUSIONS

This paper presents a methodology for traffic incident detection and prediction. The model performance is limited by the number of incident features available in real-time. However, in spite of the data limitations, the proposed approach is shown to perform significantly better than advanced spatiotemporal approaches such as TPT (Kamarianakis et al., 2012) during the early stages of an incident, after which time advanced spatiotemporal approaches can be used. Since detection of the incident is performed by the method itself via the IDM, the accuracy of the prediction can be improved by detecting incidents early, e.g. using higher frequency traffic data, such as 1-minute data, a subject for future work.

REFERENCES

B.P. Carlin, A.E. Gelfand, and A.F.M. Smith, 1992. Hierarchical Bayesian analysis of change-point problems, Applied statistics, 389–405, 1992.

Q. He, Y. Kamarianakis, K. Jintanakul, and L. Wynter, 2013. Incident duration prediction with hybrid tree-based quantile regression. S.V. Ukkusuri and K. Ozbay (eds.), Advances in Dynamic Network Modeling in Complex Transportation Systems, Complex Networks and Dynamic Systems, DOI 10.1007/978-1-4614-6243-9 12, Springer Science+Business Media, New York, 287-305.

R. Koenker, 2005. Quantile Regression. Cambridge University Press

P.T. Martin, J. Perrin, B. Hansen, R. Kump, and D. Moore, D. 2001. Incident detection algorithm evaluation. Research Report prepared for Utah Department of Transportation

W. Min and L. Wynter, 2011. Real-time road traffic prediction with spatio-temporal correlations, Transportation Research Part C: Emerging Technologies, 19(4), 606–616.

E. Parkany, and C. Xie, C. 2005. A complete review of incident detection algorithms and their deployment: what works and what doesn't. Technical report prepared for the New England Transportation Consortium, project No. 00-7.

H.J. Payne, and S.C. Tignor. 1978. Freeway Incident-Detection Algorithms Based on Decision Trees with States. Transportation Research Record: Journal of the Transportation Research Board 682: 30-37.

D. Schrank and T. Lomax, 2011. Urban Mobility Report, Texas Transportation Institute, 2011.

Y.J. Stephanedes, and A. Chassiakos, 1993. Freeway incident detection through filtering. *Transportation Research Part C*, 1: 219-233

Y. Kamarianakis, W. Shen, L. Wynter, 2012. Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO, Applied Stochastic Models in Business and Industry 28 (4), 297-315.