Analysis and real-time prediction of local incident impact on transportation networks

Yuye He, Sebastien Blandin, Laura Wynter
IBM Research Collaboratory Singapore
9 Changi Business Park Central 1
Singapore 486048
Email: heyy, sblandin, lwynter@sg.ibm.com

Barry Trager
IBM Watson Research Center
1101 Kitchawan Road
Yorktown New-York 10598
Email: bmt@us.ibm.com

***Abstract*—As an increasing proportion of the world population moves to urban areas, quality of life and economic productivity of an increasing proportion of the world population are getting impacted by traffic congestion phenomena of increasing severity. In the context of urban transportation networks of growing complexity, the management of non-recurrent road traffic events, estimated to account for close to 50% of the total travel-time delay on road networks, requires advanced predictive methods, supporting both response plan deployment and traveller information. In this work, we consider the problem of predicting how a non-recurrent traffic event, such as a road incident, impacts traffic conditions at the incident location in the future. Based on a detailed statistical analysis of a large dataset of traffic incidents, we propose a parametric piecewise-affine incident model. The model is calibrated offline using a combination of constrained and unconstrained non-linear regression methods, and is shown to provide a good fit. The prediction of local incident impact is achieved via the design of a neural network model learning the parametric fit of the incident model based on available incident features. The neural network-based prediction model is shown to outperform state-of-the-art prediction methods such as multivariate decision trees. Empirical performance of the method introduced in this work is illustrated on a large dataset of more than 50000 incidents from the city of Lyon, France, for the months of September 2013 to January 2014. Practical deployment and applicability of the proposed method in operational conditions are also discussed.**

I. Introduction

A. Motivation

According to the 2012 Urban Mobility Report [1], the cost of traffic congestion in terms of wasted time and fuel amounted to about $\$121$ billions in the US in 2011, or about 1% of the GDP of the country. With increasing urbanization pace, the societal cost of complex traffic phenomena with negative externalities is rising. However, the 2012 Urban Mobility Report [1] also highlights that a significant proportion of traffic congestion and delay incurred by commuters occurs in a non-recurrent fashion outside of peak hours.

State-of-practice intelligent transport systems are already able to estimate real-time traffic conditions and predict traffic conditions in recurrent conditions. However, the ability of the infrastructure to detect, predict, and manage non-recurrent events, for which historical data is scarce, and systemic dependencies are complex, appears as one of the key capabilities of the Smarter Transportation systems of the future, and a key to future urban development.

B. Related work

Traveller information and traffic management are two of the key capabilities of intelligent control centers. Traffic models, dating back to the 50's with the seminal work of Lighthill, Whitham and Richards [2][3], have historically been at the center of such algorithmic methods for traffic estimation, prediction and management [4][5].

With the recent explosion of mobile phones and ubiquitous traffic sensing, a large subset of the research community has focused its attention on data assimilation methods for integrating existing and new data sources into seminal traffic models, see for instance [6][7] for application of suboptimal nonlinear filtering methods derived from the Kalman filter to scalar hyperbolic conservation law models of traffic flow.

As increasing volume of data become available, purely statistical methods based either on link-level statistical learning techniques [8], or network-wide methods based on learning a correlation structure between different links [9][10], have been proposed in the recent years (see also [11] and [12] for alternative approaches). In particular, sparsity-enforcing mode-switching methods [13] have been quite successful for short and medium term prediction. Such properties of traffic data on road networks have recently been further illustrated, notably in Singapore [14].

However, given the importance of traffic incidents for transportation networks, a comparatively limited effort has been dedicated to the detection, analysis and prediction of traffic conditions resulting from non-recurrent events and incidents. Specific incident models can be found in the transportation literature [15][16], and some mathematical models of incident impact have been proposed recently [17].

The relatively rare occurrence of traffic incidents, combined with the relative sensitivity of data related to infrastructure incidents in general and the historical sparsity of urban sensor networks, explains to a large extent the limited number of data-driven methods for incident situations. It is only in the very recent years, that in-depth analysis and machine learning methods, have emerged in the literature, aiming at predicting traffic conditions resulting from incident situations, for example by analyzing incident records [18], or by mining entry logs of traffic operators [19]. Previous work on the application of decision tree model to incident records can also be found [20].

In this work, we propose to leverage a large dataset of 4

months of road traffic and incident data from the city of Lyon, France, to investigate mixed properties of road incident and traffic conditions. We then design a hybrid prediction method combining a parsimonious model-based formulation of local incident impact with an advanced nonlinear statistical learning method, in order to predict the local impact of traffic incidents.

C. Contributions

The main contributions of the work presented in this article are the following:

- Analysis of a large dataset of road traffic incidents from the city of Lyon, France, identification of key features of these incidents and quantification of local impact of these incidents on road traffic.
- Design and calibration of a local incident impact prediction model based on constrained nonlinear regression methods combined with a neural network method.
- Experimental analysis of prediction error and correlation results compared with state-of-the-art methods on real data.
- Investigation of model error structure and marginal value of data in the context of prediction of local impact of traffic incidents.

The rest of the article is organized as follows. Section II describes the available road traffic and incident datasets. In section III we introduce the statistical methods proposed in this work for estimation and prediction of incident impact, and analyze their performance in section IV. Additional insights regarding model performance and limitations are presented in section V. Section VI consists of concluding remarks and comments.

II. Data specification

In this section we present the dataset used in this work, which consists of 5 months of road traffic measurements and incident data from the city of Lyon, in France.

A. Road traffic dataset

The road traffic dataset consists of measurements of traffic occupancy, recorded at 6 mn interval for 2277 links from the road network of the inner city of Lyon, France. Traffic occupancy at a given point x represents the proportion of time a vehicle can be found at this location. The traffic occupancy ranges between 0 and 1 (or equivalently between 0 and 100%). As shown by figure 1, although few road links exhibit a very large number of incidents for the 5 month period, the average number of incidents per road link for the same period is about 47. The time evolution of these measurements for typical incidents (see section II-B for a description of incident data) is illustrated in figure 2.

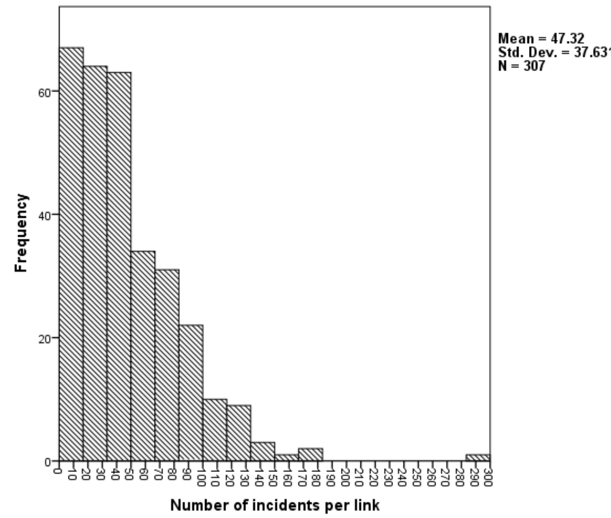


Fig. 1: Distribution of number of incidents per link.

TABLE I: Incident features.

Feature	Type
alpha	Integer
WEEKEND HOLIDAY	Binary
FUNC CLASS	Categorical
SPEED CAT	Categorical
FIRST SEVERITY	Categorical
LANES	Integer
START TIME	Time

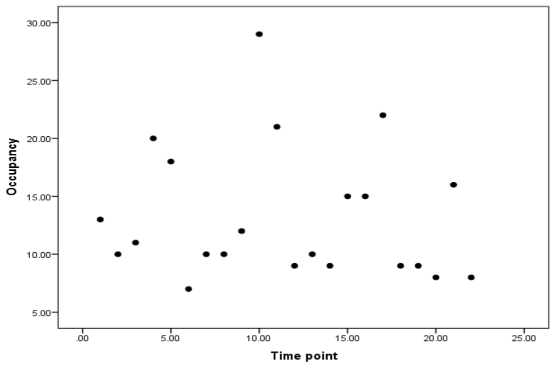
B. Incident dataset

The incident dataset consists of 55417 incidents recorded from September 1st, 2013 to January 31st, 2014. For each detected incident event, incident information including time of the incident and traffic conditions are recorded. Available incident features are presented in table I.

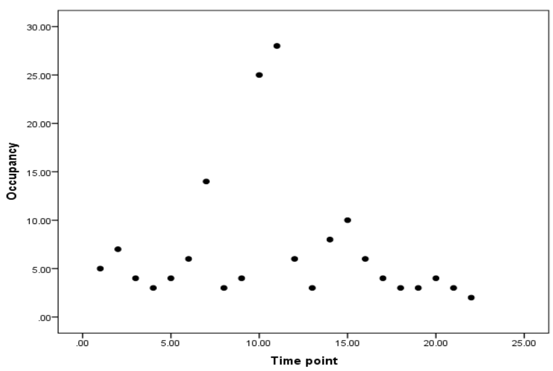
The parameter alpha denotes the value of the traffic conditions (in this case, occupancy) at the start of the incident. WEEKEND_HOLIDAY is a binary variable taking value 1 for weekends and holidays, and 0 otherwise. FUNC_CLASS is a categorical variable indicating the importance of the road for the network. SPEED_CAT is a categorical variable indicating the speed range for the road link at which the incident occurs. FIRST_SEVERITY is a categorical variable characterizing the severity of the incident at it is detected. LANES is the number of lanes. START_TIME is the detected starting time of the incident.

Based on this incident dataset (for which the end time of the incident is also available), the average incident duration is close to 3.87 time intervals, i.e., 23 minutes, as illustrated in figure 3.

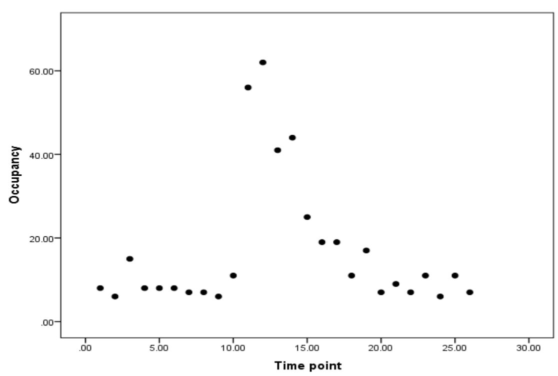
Following qualitative observations from figure 2, we propose to measure the local impact of the incident by a very general shape capturing a change in a nominal value (occupancy here), specifically; an increase in occupancy at the incident onset, possibly followed by a stationary phase, and a decrease in occupancy back to its nominal value. Analysis of the full incident dataset shows that the average value of this measure of



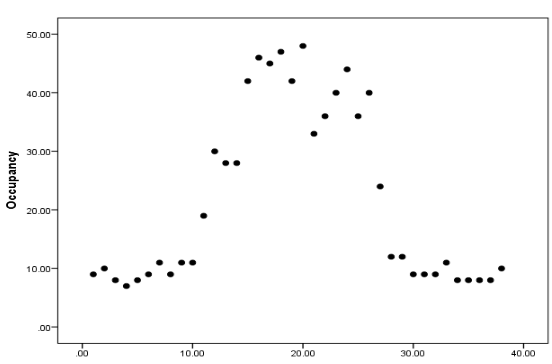
(a) Example of incident profile: highly noisy data without visible impact start and end time.



(b) Example of incident profile: significant occupancy increase around time 10 but no clear form.



(c) Example of incident profile: significant occupancy increase around time 11, gradual reduction from time 12 to time 18.



(d) Example of incident profile: significant occupancy increase from time 11 to time 27, clear form emerges.

Fig. 2: Time occupancy plot of representative incidents (time point every 6 mn)

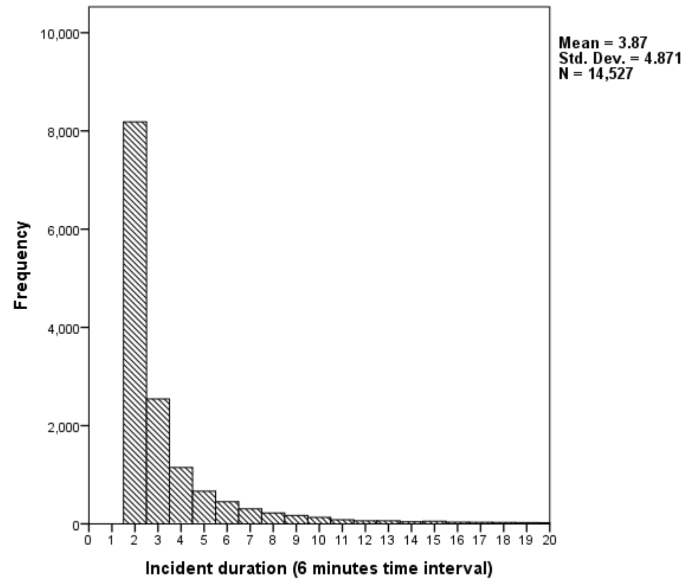


Fig. 3: Distribution of incident duration.

impact (represented by the amplitude of the occupancy change) is close to 32%, as shown in figure 4.

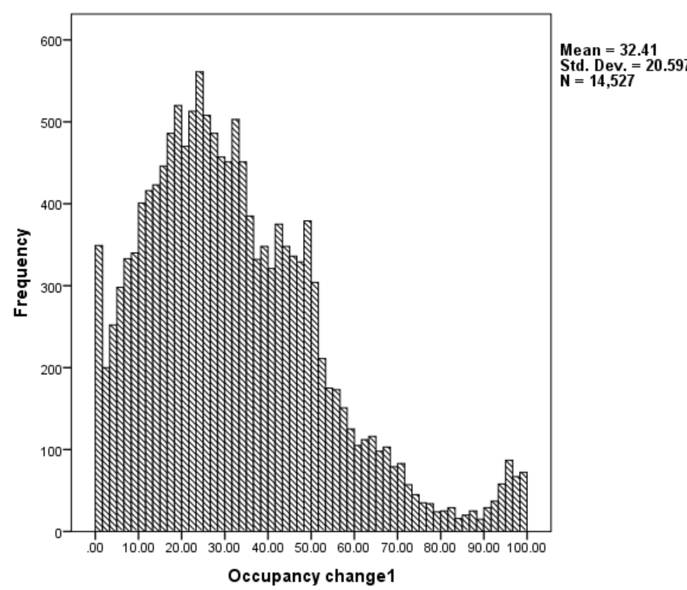


Fig. 4: Distribution of incident occupancy change.

III. Real-time prediction methods

In this section we present the prediction models introduced in this work. First we detail the regression model used at the estimation step for calibrating the model. We then present the incident model, in particular incident features and response variables, for the predictive models described in the following section.

A. Problem formulation

Based on the data analysis from previous section, we propose to build a very general model of the evolution of the local impact of the incident as a piecewise affine model. This model, while parsimonious by nature, captures the two key properties of interest for traffic management, duration of incident, as well as increase in occupancy, i.e. traffic jam.

Since different incidents may reveal different piecewise affine model, and given the relatively small size of the incident data compared to the dimensionality of the state space (we have about 50000 incidents for around 2000 locations in space and 240 locations in time per day, under 6 mn resolution data), we propose to identify incident features that are meaningful with respect to local incident impact.

With this objective in mind, we consider parameters of the piecewise affine model as a response vector for non-linear classification and regression methods, using incident features detailed in table I as predictor variable.

B. Estimation model

The calibration of the piecewise affine incident model, illustrated in figure 5, is based on a combination of constrained and unconstrained nonlinear regression models. First, erroneous records, for instance events with negative occupancy value, are filtered out using a number of automatic filters, which leaves 25823 incidents. This large number of erroneous data is fairly standard in traffic information systems. A regression model, constrained to have four change points and five segments, and with slope sign constraints, is then built for each incident. The regression model minimizes the root mean square error (RMSE) of the fit against the observed occupancy data. In order to reduce the negative effect of incidents with bad fit, only incidents with RMSE value smaller than or equal to 6 are used, which leaves 14527 incidents.

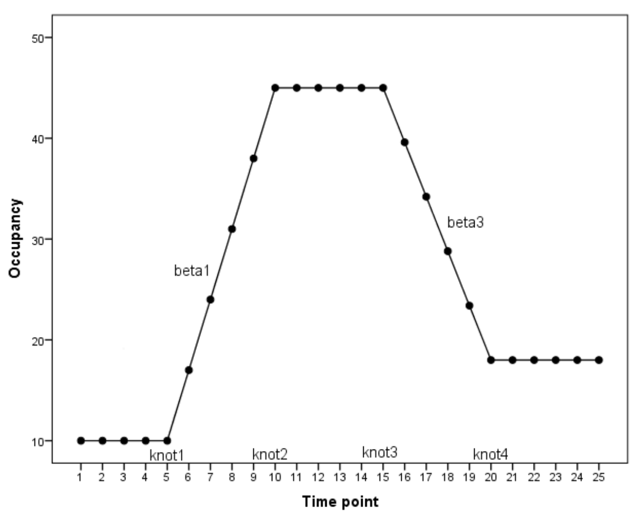


Fig. 5: A demonstration of regression model on time and occupancy data. $alpha$ is the initial occupancy; $knot_1$, $knot_2$, $knot_3$ and $knot_4$ are the four change points; $beta_1$ and $beta_3$ are the slopes.

The location of the change points and slopes of non-horizontal segments are collected as parameters for the regression model. The estimated parameters together with the incident features provide the input for the subsequent incident prediction phase.

C. Local incident impact parametrization

We parameterize the local incident impact with the non-linear regression model parameters from the estimation model presented in the previous section.

We denote k_{21}, k_{31} and k_{41} the duration between change points $knot_2$ and $knot_1$, $knot_3$ and $knot_1$, $knot_4$ and $knot_1$, respectively. The variables occu_change$_1$ and occu_change$_2$ are used instead of $beta_1$ and $beta_3$ to denote the absolute occupancy change between the first and second horizontal phase, the second and the third horizontal phase, respectively.

These notations are illustrated in figure 5, and a summary of predictor and response variables, for both univariate and multivariate methods, can be found in table II.

TABLE II: Variable information for three types of models.

Predictor variables	Alpha, WEEKEND HOLIDAY, FUNC CLASS, SPEED CAT, FIRST SEVERITY, LANES, START TIME
Response variables for univariate model	k_{41}
Response variables for multivariate model	$knot_1$, k_{21}, k_{31}, k_{41}, occu_change$_1$ and occu_change$_2$

D. Prediction model

For predicting local incident impact based on incident features, we compare the three following methods: univariate decision tree (UVDT), multivariate decision tree (MVDT) [21] and neural network (NN) method. The univariate decision tree method uses only one response variable to build the decision tree and predicts the local incident impact by taking the average of the six regression model parameters falling into the same node as the final prediction value. Multivariate decision tree and neural network methods use all six parameters as response variables. The univariate decision tree and neural network models are built using SPSS statistics 22.0 and the multivariate decision tree model is built using the *mvpart* package from statistical software R. For each model, 80% of the data set is used as training set and and 20% as validation set. The model is trained on the training set using 5-fold cross validation (CV) and the final model is applied on the validation set (external validation). The predicted parameters are used in the piecewise affine regression model to calculate the predicted incident occupancy values and assess performance against observed occupancy values.

1) Univariate decision tree: Univariate decision tree is a popular predictive method for classification and regression applications. We use the difference between $knot_4$ and $knot_1$, noted k_{41}, representing the estimated incident duration, as response variable. We then assume that incidents with similar duration have similar local impact, and use the average of each regression parameters within the same node of the decision tree for prediction. In this work, the SPSS statistics 22.0 implementation of the Chi-squared Automatic Interaction

Detector (CHAID) method is used for its better performance and eaiser interpretation.

2) Multivariate decision tree: Multivariate decision tree generally outperform univariate decision trees because sample instances can be split according to more than a single response variable at each internal node [21]. In this work, we use as response variables the entire set of regression parameters from the local incident impact model, as detailed in table II. The method is developed by extending the R implementation of the Classification And Regression Tree (CART) algorithm [22].

3) Neural network: Neural network methods have been widely used in learning applications with multiple response variables. The network consists of one input layer, one or more hidden layers and one output layer, each consisting of several neurons. Each neuron processes its inputs and generates one output value that is passed to the neurons in the next layer. Each neuron in the input layer delivers the value of one predictor. In each layer, the weights are usually initialized with some small random value. The training process works by adjusting the weights to minimize the error of the neural network (NN) output against the actual output.

There are two NN modeling methods available in SPSS statistics 22.0 implementation. The multilayer perceptron (MLP) method is preferred over the radial basis function (RBF) method, because MLP allows multiple hidden layers and generally is more accurate than RBF method. The most popular algorithm for training MLP NN model is the back-propagation (BP) algorithm.

For this algorithm, the error computed from the output layer is back-propagated through the network, and the weights are adjusted according to their contribution to the error function. Essentially, back-propagation performs a local gradient search, and hence it does not guarantee to reach the global minimum. For each individual, weights are adjusted to minimize the error computed from the output layer. After the network reaches a convergence or satisfying certain stopping criteria, the relation between the predictors and the response variables is returned. The information for the NN model built for this work is summarized in table III.

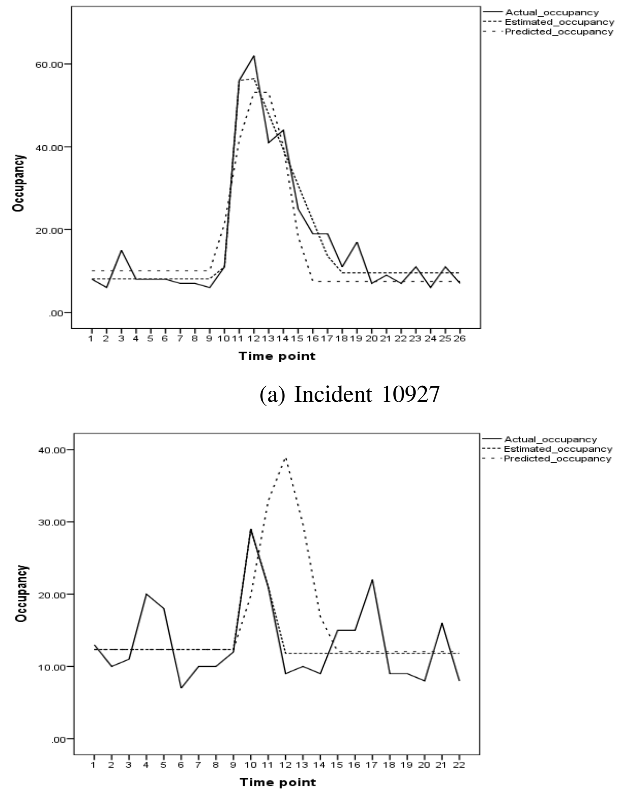
IV. Experimental results

In this section we present the performance results of the method introduced in this work and detailed in the previous section.

A. Qualitative analysis

To have a general idea of the effect of incident estimation and prediction process on observed incident data, two sample incidents with observed, estimated and predicted occupancy are shown in figure 6.

A good qualitative match can be observed between the observed, estimated and predicted occupancy pattern for incident 10927. However, there is a big deviation between estimated and predicted occupancy for incident 184437, which indicates the prediction is not very accurate. In the following section, we present a more comprehensive and quantitative view of the performance of the method on the entire dataset.



(a) Incident 10927

(b) Incident 184437

Fig. 6: Time occupancy plot of sample incidents

B. Prediction model performance

We present results on the model performance for two error metrics. First we present results for correlation between estimated piecewise affine local incident impact model and predicted piecewise affine local incident impact model. We then illustrate how the method introduced in this work performs in terms of mean absolute prediction error.

1) Correlation results: The correlation results for cross validation and external validation for three types of models are presented in Table IV.

Correlation results show the methods presented perform much better for variables directly related to incident impact (given the parsimonious nature of the model, occupancy variables occu_change$_1$ and occu_change$_2$ provide a direct measure of the severity of the incident), whereas for indirect impact variables, i.e. time related parameters $knot_1$, k_{21}, k_{31}, k_{41}, the correlation is lower. This may be explained by the fact that occupancy variables are state variables constitutive of traffic conditions, whereas change point time are not as directly related to traffic conditions, and more subject to exogenous dependency and variations. Additionally one may note that the time resolution of the data (6 minutes) is relatively large compared to the typical duration of each of the phase of the incident.

Comparative analysis of the results of the univariate and multivariate methods show that the multivariate methods

TABLE III: Information for NN model.

Network Information			
Input Layer	Factors	1	FUNC_CLASS
		2	SPEED_CAT
		3	WEEKEND_HOLIDAY
	Covariates	1	$alpha$
		2	START_TIME
		3	FIRST_SEVERITY
		4	LANES
	Number of Units	14	
	Rescaling Method for Covariates	Standardized	
Hidden Layer(s)	Number of Hidden Layers	2	
	Number of Units in Hidden Layer 1a	10	
	Number of Units in Hidden Layer 2a	8	
	Activation Function	Hyperbolic tangent	
Output Layer	Dependent Variables	1	$knot_1$
		2	k_{21}
		3	k_{31}
		4	k_{41}
		5	occu_change$_1$
		6	occu_change$_2$
	Number of Units	6	
	Rescaling Method for Scale Dependents	Standardized	
	Activation Function	Identity	
	Error Function	Sum of Squares	

TABLE IV: Correlation results for neural network model (NN), multivariate decision tree model (MVDT), and univariate model with k_{41} as response variable (DT_K41).

			knot$_1$	**k$_{21}$**	**k$_{31}$**	**k$_{41}$**	**occu_change$_1$**	**occu_change$_2$**
Training performance	**NN**	**5-fold CV**	0.156	0.150	0.220	0.248	0.549	0.438
		External validation	0.195	0.168	0.255	0.279	0.570	0.469
	MVDT	**5-fold CV**	0.150	0.145	0.139	0.157	0.570	0.478
		External validation	0.154	0.110	0.112	0.124	0.571	0.451
	DT_K41	**5-fold CV**	0.128	0.150	0.246	0.282	0.284	0.255
		External validation	0.136	0.150	0.248	0.282	0.299	0.281
Validation performance	**NN**	**5-fold CV**	0.151	0.151	0.215	0.245	0.547	0.434
		External validation	0.210	0.151	0.227	0.230	0.591	0.477
	MVDT	**5-fold CV**	0.132	0.123	0.112	0.133	0.537	0.436
		External validation	0.154	0.110	0.112	0.124	0.571	0.451
	DT_K41	**5-fold CV**	0.112	0.122	0.218	0.256	0.279	0.265
		External validation	0.063	0.009	0.107	0.123	0.202	0.220

(MVDT and NN) outperform the univariate method DT_K41 consisting of a decision tree method using the estimated incident duration as response. This illustrates that different piecewise affine incident model parameters exhibit independent statistics, and while the results of DT_K41 are competitive for the estimated incident duration, the calibrated univariate decision tree is not able to reliably estimate other model parameters.

The correlation results also illustrate that NN outperforms MVDT for all response variables, although by a relatively slight margin. Additionally NN model is used in subsequent application of this method because it can capture more complex relationship than MVDT.

2) Mean absolute error: For each incident, based on the predicted set of parameters analyzed in previous section, a piecewise affine regression model is constructed for each incident and the occupancy value for each time point is predicted (as illustrated in figure 6). We assess the model performance by comparing the observed occupancy value with the predicted occupancy value. We use the mean absolute error (MAE) as error metric.

Given that different incidents may have quite different durations (see for instance figure 3), we propose to display the prediction error as a function of the normalized prediction horizon. Hence although we make the prediction in the standard time space, before displaying the results, we dilate (for short incidents) or contract (for long incidents) the time axis in order to align the location of each change point in the final display of results. All outputs are normalized before display on the following value for the change points; $knot_1$=5, $knot_2$=10, $knot_3$=15 and $knot_4$=20.

The distribution of absolute error (AE) between actual and predicted occupancy for all incidents is shown in figure 7. Specifically, the box-plot at a given (normalized) time is generated by considering the occupancy prediction for each of the incident, for this (normalized) time. Results indicate that the NN model achieves the smallest prediction error. The prediction error for MVDT model is slightly higher than NN model while the DT_K41 model gives the highest prediction error. This is consistent with the correlation result shown in Table IV.

The average value of the absolute error is shown in figure 8, illustrating the comparatively better performance of the NN model. Given the actual shape of evolution of the occupancy

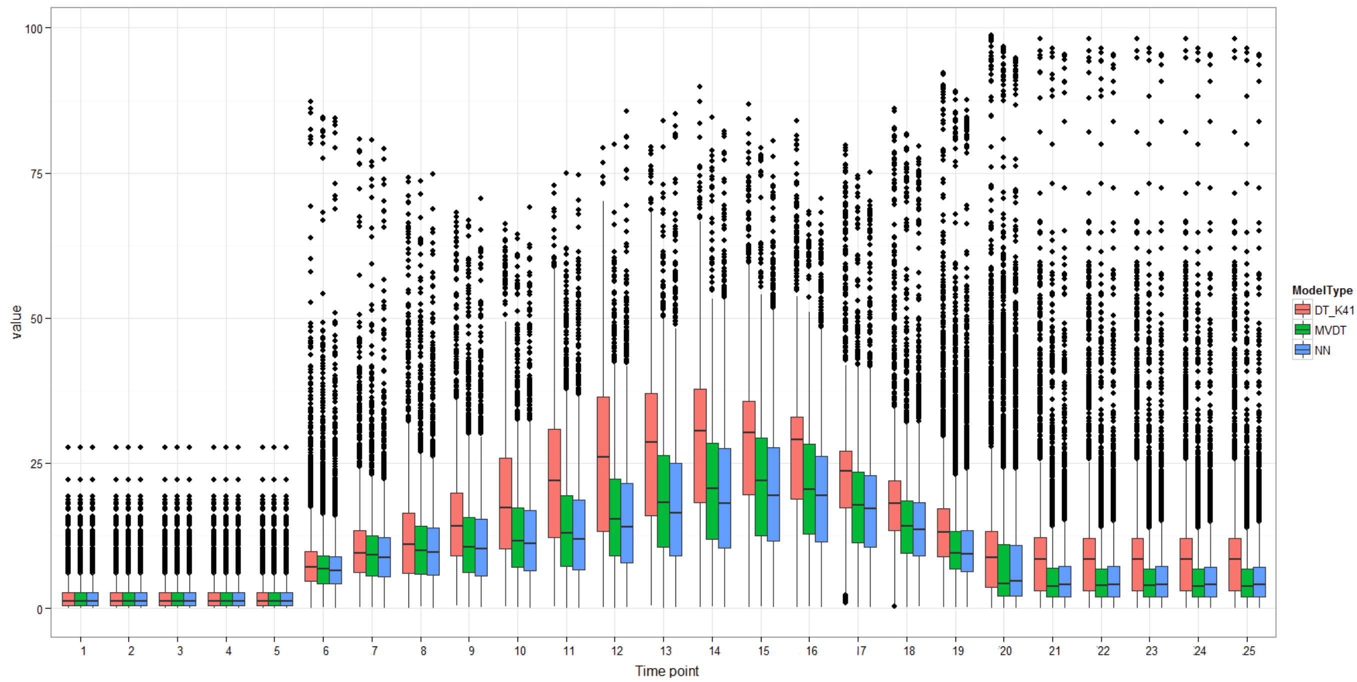


Fig. 7: Distribution of absolute error between actual and predicted occupancy value for three models.

during an incident, it is also clear that the prediction error increases as the incident evolves, and starts decreasing only when traffic conditions return to their nominal value. The maximal MAE value is reached around time step 15, close to the third change point, which corresponds to the beginning of the last phase. One may note that this is the time at which the incident is at its peak and the observed occupancy is maximal, empirically observed to reach values close to 60 on average.

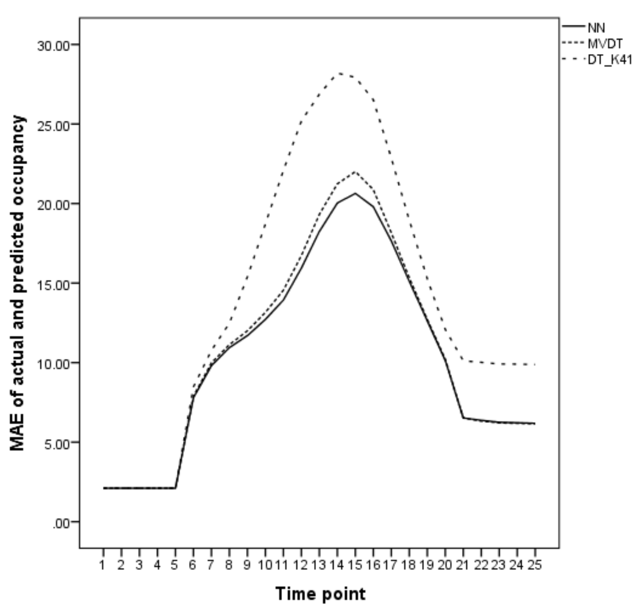


Fig. 8: MAE of actual and predicted occupancy value for three models.

V. Additional Insights into Incident Predictability

In this section we illustrate how additional properties of the method introduced in this work for prediction of local incident impact, can be leveraged for further analysis of road incident and traffic conditions by command center operators. We first present results on the relative classification power of incident features, informing practitioners about the value of data in this context, and then focus on classifying the incidents according to their prediction accuracy, in order to selectively invoke the system built around this method, depending on the targeted accuracy.

A. Predictor analysis for neural network model

In order to assess the relative predictor importance for the neural network model, we propose to use the trained neural network model to test the sensitivity of the response variable to the changes in each predictor. This is achieved by varying the values of each predictor while keeping all the other predictors at their respective means [23]. The variation of output with respect to the variation of each predictor is recorded and a resulting normalized measure of the relative importance for each predictor is produced. The value 0 corresponds to no effect on the prediction while a value 1 corresponds to a predictor determining the prediction completely).

In figure 9, we display the normalized importance for each predictor for the neural network model. The top two important variables for the model are the initial occupancy before the incident (noted $alpha$) and the number of lanes involved (LANES). This result is consistent with previous studies for incident duration prediction [24][25].

The time and date of the incident are not that important for the predictive model which may be due to the fact that incidents used in this work are by definition non-recurrent events so the time information does not hold any specific information. Additionally, we are here mostly concerned with the local impact of the incident, and we are not studying how much congestion is created in a sub-network around due to the incident, for which time of day (peak hour or non-peak hour in particular) would have some predictive power even for non-recurrent events.

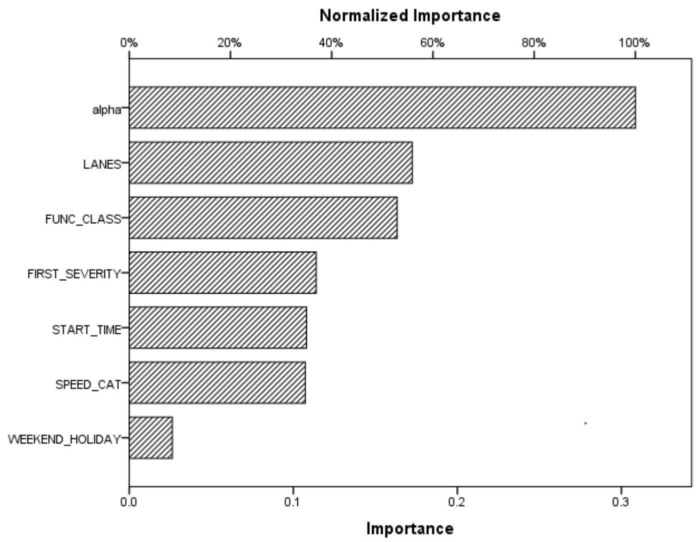


Fig. 9: Predictor importance for neural network model.

B. Incident impact predictability using binary classification

In this section we analyze the extent to which we can predict the accuracy of the prediction of the proposed method based on incident features. This would allow selection of a target quality threshold by the user before using the system.

In order to achieve this predictability result, we propose to use a binary classification method. First we remove from the entire set of 28536 incidents, the incidents for which the regression model analysis returns negative $alpha$ or sign error. The filtered dataset consists of 26476 incidents.

In order to determine the optimal cutoff threshold, we then apply a binary classification process, with the RMSE for the regression model, considered a response variable. The mean of the RMSE value for all incidents is around 5.5 so seven cutoff values 3, 4, 5, 5.5, 6, 7 and 8 are selected. For each cutoff value, the incidents are separated into two classes; incidents with RMSE smaller than or equal to cutoff value make the positive class (good fit) while incidents with RMSE greater than cutoff value constitute the negative class (bad fit).

A binary classification model is calibrated using decision tree with 5-fold cross validation. The performance of the classification model for cross validation is recorded in table V.

The result shows that when the cutoff value increases, the sensitivity SE increases and the specificity SP decreases. When

TABLE V: Prediction result of binary classification model. SE is sensitivity, SP is specificity and ACC is total accuracy.

RMSE cutoff	SE(%)	SP(%)	ACC(%)
3	40.9	87.5	75.3
4	49	81.5	69.3
5	68.7	67.4	68
5.5	77.3	59.5	68.9
6	85	50.5	70.4
7	89.9	39.7	73.1
8	88.6	43.5	76.5

cutoff value equals 6, both SE and SP values are greater than 50%, which means that the binary model generated with cutoff value 6 is able to separate the positive and negative classes, i.e., good and bad fits for the regression model, better than a random guess, and this cutoff value also dominates the other cutoff value.

VI. Conclusion

In this work, we develop a new method for detailed prediction of local impact of road incidents on traffic conditions. Based on analysis of empirical incident impact on traffic occupancy, we design a local incident impact model, and calibrate the model using a constrained non-linear regression method which provides good results. Prediction of the local incident impact is achieved using a neural network model, shown to outperform other state-of-the-art method.

The major limitation of this work, apparent in the illustrated correlation results, likes is the relatively small amount of data available to characterize the incidents. Only seven incident features are available to develop the model which may not be sufficient to fully capture an incident pattern.

It has been reported the most important variables for incident duration prediction include number of lanes affected, vehicles' information, time of day, police response time and weather condition [24][25].

A second limitation of this work includes the low quality of the incident data. Given the fact that the incidents were provided by a black-box method combining operator expertise with statistical learning, there is no intrinsic consistency between the detection method and the prediction technique.

As illustrated by the numerical results, under the data limitations mentioned, the method introduced in this work, consisting of feeding a neural network prediction model with incident features for parametric incident impact treated as a response, is shown to outperform decision-tree based method, and to provide an informative prediction of the local impact of road incident on traffic condition. In particular, prediction of the incident duration and the increase of traffic occupancy, which are key quantities for the operator managing the response to traffic flow, are relatively well predicted. Extensions to this work include the combination of the proposed approach with a Bayesian change-point method for online update of the prediction as new data streams into the system, as well as the development of a spatial impact propagation method for the incident, in order to provide operators and commuters with a prediction of non-local incident impact.

References

[1] D. Schrank, B. Eisele, and T. Lomax, "The 2012 urban mobility report," 2012.

[2] M. Lighthill and G. Whitham, "On kinematic waves, ii : a theory of traffic flow on long crowded roads," *Proceedings of the Royal Society of London*, vol. 229, no. 1178, pp. 317–345, 1956.

[3] P. Richards, "Shock waves on the highway," *Operations Research*, vol. 4, no. 1, pp. 42–51, 1956.

[4] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended kalman filter: a general approach," *Transportation Research Part B*, vol. 39, no. 2, pp. 141–167, 2005.

[5] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and Y. Wang, "Review of road traffic control strategies," *Proceedings of the IEEE*, vol. 91, no. 12, pp. 2043–2067, 2003.

[6] S. Blandin, A. Couque, A. Bayen, and D. Work, "On sequential data assimilation for scalar macroscopic traffic flow models," *Physica D*, vol. 241, pp. 1421–1440, 2012, doi:10.1016/j.physd.2012.05.005.

[7] D. B. Work, S. Blandin, O.-P. Tossavainen, B. Piccoli, and A. M. Bayen, "A traffic model for velocity data assimilation," *Applied Mathematics Research eXpress*, vol. 2010, no. 1, pp. 1–35, 2010.

[8] M. Asif, J. Dauwels, C. Y. Goh, A. Oran, E. Fathi, M. Xu, M. Dhanya, N. Mitrovic, and P. Jaillet, "Unsupervised learning based performance analysis of n-support vector regression for speed prediction of a large road network," in *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*. IEEE, 2012, pp. 983–988.

[9] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 606–616, 2011.

[10] S. Samaranayake, S. Blandin, and A. Bayen, "Learning the dependency structure of highway networks for traffic forecast," in *IEEE Conference on Decision and Control*, Orlando, FL, December 2011, doi:10.1109/CDC.2008.4739016.

[11] J. He, Q. He, G. Swirszcz, Y. Kamarianakis, R. Lawrence, W. Shen, and L. Wynter, "Ensemble-based method for task 2: Predicting traffic jam." in *ICDM Workshops*, 2010, pp. 1363–1365.

[12] W. Shen, Y. Kamarianakis, L. Wynter, J. He, Q. He, R. Lawrence, and G. Swirszcz, "Traffic velocity prediction using gps data: Ieee icdm contest task 3 report," in *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1369–1371.

[13] Y. Kamarianakis, W. Shen, and L. Wynter, "Real-time road traffic forecasting using regime-switching space-time models and adaptive lasso," *Applied Stochastic Models in Business and Industry*, vol. 28, no. 4, pp. 297–315, 2012.

[14] M. T. Asif, S. Kannan, J. Dauwels, and P. Jaillet, "Data compression techniques for urban traffic data," in *Computational Intelligence in Vehicles and Transportation Systems (CIVTS), 2013 IEEE Symposium*. IEEE, 2013, pp. 44–49.

[15] L. Qin and B. L. Smith, "Characterization of accident capacity reduction," University of Virginia, Tech. Rep., 2001.

[16] V. L. Knoop, S. P. Hoogendoorn, and H. J. Van Zuylen, "Capacity reduction at incidents: empirical data collected from a helicopter," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2071, no. 1, pp. 19–25, 2008.

[17] J.-B. Sheu, Y.-H. Chou, and A. Chen, "Stochastic modeling and real-time prediction of incident effects on surface street traffic congestion," *Applied Mathematical Modelling*, vol. 28, no. 5, pp. 445–468, 2004.

[18] F. C. Pereira, O. Lederman, and M. Ben-Akiva, "Approximating incident occurrence time with a change-point latent variable framework," in *Transportation Research Board 93rd Annual Meeting*, no. 14-5338, 2014.

[19] F. C. Pereira, F. Rodrigues, and M. Ben-Akiva, "Text analysis in incident duration prediction," *Transportation Research Part C: Emerging Technologies*, vol. 37, pp. 177–192, 2013.

[20] S. Blandin, V. Gopal, K. Thirumalai, and J. Cheong, "Prediction of duration and impact of non-recurrent events in transportation networks," TRISTAN VIII, San Pedro de Atacama, Chile, June 9th - 14th 2013.

[21] C. Brodley and P. Utgoff, "Multivariate decision trees," *Machine Learning*, vol. 19, no. 1, pp. 45–77, 1995.

[22] G. DeAth, "mvpart: Multivariate partitioning," 2013.

[23] D. Delen, R. Sharda, and M. Bessonov, "Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks," *Accident Analysis & Prevention*, vol. 38, no. 3, pp. 434–444, 2006.

[24] A. Garib, A. Radwan, and H. Al-Deek, "Estimating magnitude and duration of incident delays," *Journal of Transportation Engineering*, vol. 123, no. 6, pp. 459–466, 1997.

[25] H.-L. Chang and T.-P. Chang, "Prediction of freeway incident duration based on classification tree analysis," *Journal of the Eastern Asia Society for Transportation Studies*, vol. 10, pp. 1964–1977, 2013.