Anomaly-Aware Traffic Prediction Based on Automated Conditional Information Fusion

1st Bin Sun
Department of Creative Technologies
Blekinge Institute of Technology
Karlskrona 37179, Sweden
bin.sun@bth.se

2nd Wei Cheng
Kunming University of Science and Technology
Kunming 650093, China
Blekinge Institute of Technology
Karlskrona 37179, Sweden
Corresponding Author: wei.cheng@bth.se

3rd Liyao Ma
School of Electrical Engineering
University of Jinan
Jinan 250022, China
cse_maly@ujn.edu.cn

4th Prashant Goswami
Department of Creative Technologies
Blekinge Institute of Technology
Karlskrona 37179, Sweden
prashant.goswami@bth.se

***Abstract*—Reliable and accurate short-term traffic prediction plays a key role in modern intelligent transportation systems (ITS) for achieving efficient traffic management and accident detection. Previous work has investigated this topic but lacks study on automated anomaly detection and conditional information fusion for ensemble methods. This works aims to improve prediction accuracy by fusing information considering different traffic conditions in ensemble methods. In addition to conditional information fusion, a day-week decomposition (DWD) method is introduced for preprocessing before anomaly detection. A *k*-nearest neighbours (*k*NN) based ensemble method is used as an example. Real-world data are used to test the proposed method with stratified ten-fold cross-validation. The results show that the proposed method with incident labels improves predictions up to 15.3% and the DWD enhanced anomaly-detection improves predictions up to 8.96%. Conditional information fusion improves ensemble prediction methods, especially for incident traffic. The proposed method works well with enhanced detections and the procedure is fully automated. The accurate predictions lead to more robust traffic control and routing systems.**

***Keywords*—Intelligent Transportation Systems (ITS), Short-Term Traffic Prediction, Information Fusion, Time Series Decomposition, *k*-Nearest Neighbours (*k*NN)**

I. Introduction

Intelligent transportation systems (ITS) are becoming more and more effective to avoid or resolve severe congestions and accidents [1], [2]. Reliable and accurate short-term traffic prediction is fundamental for modern ITS to achieve this target [3]. As a complex task, it has been studied in the past few decades using different schemes [4].

One widely used scheme is to fuse information according to different conditions to improve prediction accuracy [5]–[8], However, finding the way to distinguish the traffic conditions is a problem as the related records for data labelling are messy. For messy labels or totally non-labelled data, some human resources may be needed such as labelling [9] or active learning [10]. In addition, there are different ways to categorize traffic conditions from either mathematical aspect or traffic engineering aspect and even more detailed aspects [11]–[13]. Normal vs. abnormal conditions are more suitable for a higher level analysis [14]. Though some previous methods are anomaly-aware, they are not automated ensemble methods and require manual parameter tuning, which are detailed in the next section. This work proposes the idea to predict traffic with the awareness of incidents or anomalies to enhance automated ensemble method. Besides, this work solves the issue of dynamic traffic fluctuation and multi-seasonality by introducing a day-week model for preprocessing before anomaly detection.

II. Related Work

A. Conditional Prediction

It has become consensus that different prediction methods or parameters should be used for different traffic conditions to improve accuracy, especially during extreme events [5]. For example, prediction-after-classification approaches [8] has been used, but it is considering one whole day as one data point without dynamic time series characteristics.

Many studies focus on choosing training data or nearest neighbours by clustering. For example, one work [6] found four different pattern changes, awareness of related patterns improves prediction accuracy. Theodorou et al. [7] considered typical vs. atypical conditions, but seasonal autoregressive integrated moving average (SARIMA) is their base method which is not accurate even with automated parameter tuning [15]. A flow-aware ensemble method weighted parameter tuples (WPT) is proposed targeting dynamic flow characteristics [15]. It has been compared with traditional time-aware predictions [16] and performs well. Ensemble methods have shown the ability to fuse information to improve prediction accuracy [17], [18] and to avoid manual parameter tuning such as WPT.

However, we have not seen any conditional ensemble methods with automated parameter tuning for short-term traffic prediction considering incidents or anomalies. Thus, based on the automated ensemble method WPT, this work proposes the idea to predict traffic with the awareness of incidents or anomalies.

B. Anomaly Detection

One group of anomaly detection methods are based on distribution, distance or density. This group includes traditional distribution based methods like dynamic Poisson distribution based detection [19], Chi-square test [20], and modern machine learning methods such as local outlier factor [21], [22], one class support vector machine [23] and iForest [24]. However, distance/density-based methods require the time series should be transformed into points in a multidimensional space. This transformation loses the time series characteristics. Dynamic Poisson distribution uses more time-domain information, but it only detects local outliers.

Another way is to use regression methods [25] to build a model to get residuals and detect outliers according to a fixed or dynamic threshold. SARIMA [26] is widely used to build time series models. However, it is hard to automatically find suitable SARIMA parameters, and the performance is worse than kNN [15], [27]. Some automated versions of SARIMA [28] simply found “no suitable ARIMA model” for complex traffic data.

To get a better seasonality estimation, seasonal and trend decomposition with Loess (STL)-based methods are frequently used [29], [30]. One detection algorithm *TSoutliers* [29] uses STL and super smoother to build time series model and uses student-distribution based quantiles to detect outliers. This will detect additive outliers or pulses. Another recent algorithm is *SH-ESD* which is introduced by Twitter [31]. The basic algorithm is extreme studentized deviate test (ESD test) which is proposed by Roster [32]. The trend and seasonality should be removed before ESD. To remove it, seasonal ESD (S-ESD) [31] is using an improved version of STL [30] for time series decomposition. By using robust statistics such as median and median absolute deviation, S-ESD is improved to seasonal hybrid ESD (SH-ESD). SH-ESD has shown promising results [33]–[35] and is used as a base benchmark in this work together with TSoutliers.

For those regression-detection algorithms, one issue is that the dynamic traffic fluctuation is not considered. With higher traffic, comes higher fluctuation and deviation. Another issue is that the multi-seasonality characteristics in traffic time series are not considered. Thus, we propose to build a model for traffic by considering dynamic traffic fluctuation and multi-seasonality.

III. Methodology

The proposed method improves ensemble kNN by conditional information fusion and enhanced anomaly detection.

A. kNN Regression for Prediction

The basic kNN regression algorithm for prediction is shown in Fig. 1. Suppose there is a traffic time series on time ticks: $\cdots, t-2, t-1, t, \cdots$ and we need to predict traffic flow at time t. The time interval between two neighbouring time ticks ranges from 1 minute to 15 minutes depending on different systems.

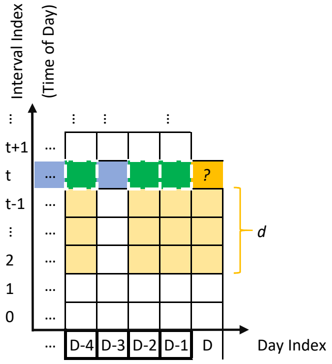


Fig. 1. Prediction for time t using basic kNN regression without window. History data last until time point $t-1$. Search step length (lag) is d.

To predict the traffic at time t, two main steps are conducted. First, a state vector is constructed using the most recent data as a new query to represent current traffic state.

$$\mathbf{S}_{[t]} = \begin{bmatrix} r_{t-1} & r_{t-2} & \cdots & r_{t-d} \\ s_{t-1} & s_{t-2} & \cdots & s_{t-d} \end{bmatrix} \tag{1}$$

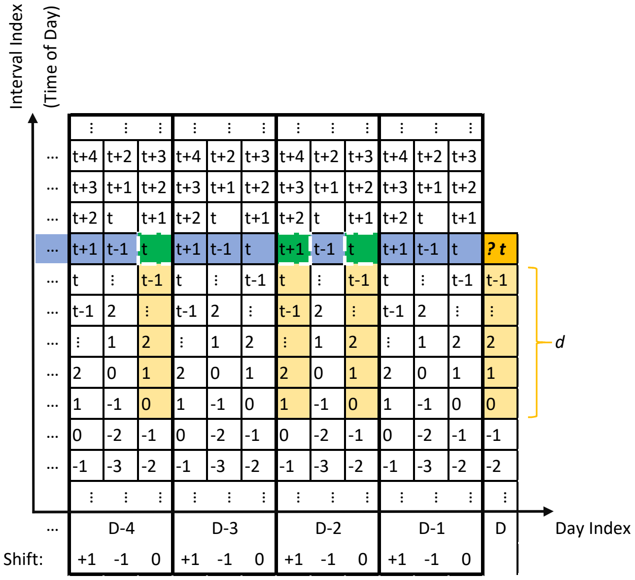
where r is flow rate and s is speed, d is search step length (lag), which means d data points from $t-1$ until $t-d$ are used. Second, find the state vectors (nearest neighbours) from history data that are similar to current query state vector. For instance, in Fig. 1, state vectors from Day_{-1} (yesterday), Day_{-2} and Day_{-4} are the three nearest neighbours. Thus, the history values at time t of those three days are considered as neighbours' predictions and are averaged as the final prediction of traffic at time t of today.

This procedure has been improved by introducing window size (v), which is the maximum time shift during neighbour searching. If $v = 1$, all possible shifts are -1,0,+1. For instance, Fig. 2 shows $v = 1$ and the number of neighbours within each day increased to three. The last step is still averaging the values of all k nearest neighbours' predictions as the final prediction.

Now, there are three parameters in kNN regression to tune. The aforementioned method, WPT, is able to consider all three parameters at the same time and predict traffic automatically without manual parameter tuning. WPT is aware of the flow statistics, but not the flow conditions. The following content adds condition awareness to WPT and enhances the condition detection algorithms.

B. Conditional Information Fusion for kNN

As mentioned in Section II, WPT fuses information according to different flow rates. We modified it to fuse the information according to the flow conditions, i.e. normal vs. abnormal data, by not separating the data for different flow rates. This replaces the flow-awareness by anomaly-awareness. Abnormal history time points and normal data points are considered as two different groups. The time points in those



	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
...	t+4	t+2	t+3	t+4	t+2	t+3	t+4	t+2	t+3	t+4	t+2	t+3	
...	t+3	t+1	t+2	t+3	t+1	t+2	t+3	t+1	t+2	t+3	t+1	t+2	
...	t+2	t	t+1	t+2	t	t+1	t+2	t	t+1	t+2	t	t+1	
...	t+1	t-1	t	t+1	t-1	t	t+1	t-1	t	t+1	t-1	t	? t
...	t	⋮	t-1	t	⋮	t-1	t	⋮	t-1	t	⋮	t-1	t-1
...	t-1	2	⋮	t-1	2	⋮	t-1	2	⋮	t-1	2	⋮	⋮
...	⋮	1	2	⋮	1	2	⋮	1	2	⋮	1	2	2
...	2	0	1	2	0	1	2	0	1	2	0	1	1
...	1	-1	0	1	-1	0	1	-1	0	1	-1	0	0
...	0	-2	-1	0	-2	-1	0	-2	-1	0	-2	-1	-1
...	-1	-3	-2	-1	-3	-2	-1	-3	-2	-1	-3	-2	-2
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
...	D-4			D-3			D-2			D-1			D
Shift:	+1	-1	0	+1	-1	0	+1	-1	0	+1	-1	0	

Fig. 2. Prediction for time t using kNN with window size $v = 1$ (shifts: -1,0,+1). History data last until time point $t - 1$. Search step length (lag) is d.

two groups are used separately to train parameter tuples. Later, for the prediction of traffic flow, one of the two groups trained parameter tuples weights will be used respectively. Thus, the prediction is aware of normal vs. abnormal traffic.

C. DWD Preprocessing before Anomaly Detection

The purpose of preprocessing is to get normalized residuals to feed anomaly detection algorithms. The daily and weekly seasonality plays a key role in traffic engineering, so we propose a method named day-week-decomposition (DWD) to get rid of the daily and weekly seasonality. DWD will produce a *day-week model* which contains seven *day-of-week submodels*. The traffic flow data are averaged by a 15 min window to build the day-week model as it is the minimum time to get stable traffic flow [36]. To build the submodels, the data time points are divided into seven subsets according to their days of week. The median values of each time-of-day are used as one time point in a specific submodel. For example, if the time interval is 5 minutes, one day includes 288 time points. Each time point is the median value of the same time of day from one whole subset.

The basic idea to calculate residuals is subtracting the modelled values from the real values. Though for the same day of week, the traffic patterns are similar, the flow levels are different. Thus, for each day, the corresponding submodel is *scaled* and *shifted* to the day's data, hence *moved-submodels*. This transformation is done by robust fitting of nonlinear regression (*nlrob*) [37]. Consecutive zeros in this dataset are removed as they are from device malfunctioning or road closing and makes *nlrob* to fail. Then the residuals are calculated via subtracting the moved day-of-week submodels from the real values. Additionally, to smooth connections between neighbouring days during the scaling and shifting, the median occurring time of all the lowest points of the days are used as the beginnings and endings of each day. As higher flow rates produce higher fluctuations, the residuals are divided by their related standard deviations considering similar flow rates. As the residuals are derived from the data before 15 minutes smoothing, the residuals are smoothed. Finally, we get the normalized residuals.

Later, both the original data and the DWD generated data are sent to TSoutliers or SH-ESD to detect anomalies. Thus, the data are divided into two groups, normal data and abnormal data.

IV. Experiments

A. Data Specification

The real world data are collected during April 2013 - May 2014 from a highway named Kunshi [9]. One device sends a monitored flow record at five-minute intervals. Each record contains some traffic statistics such as flow rate and average speed. This road carries under-saturated flow except in holidays noons.

Ground truth incident labels are generated by using the extended system mentioned in [9]. The data are imputed using the method from [38] before any processing.

B. Experimental Setup

The ensemble procedure is using the parameter generating rules from [39]. They developed an algorithm which is time-aware while this work develops an algorithm which is anomaly-aware. The values of k, d and v for the basic kNN are as follows: $\mathbb{K} = \{2, 4, 8, \cdots, 256\}$, $\mathbb{D} = \{2, 4, 8, \cdots, 256\}$, $\mathbb{V} = \{0, 4, 8, 16, 32\}$.

As the ensemble method consumes much time to run on a normal central processing unit (CPU), graphics processing unit (GPU) computing is used. The basic kNN experiments are conducted using CUDA [40] driver v8.0 and runtime v7.5 on a Nvidia Titan X Pascal. To get rid of the influence of missing value imputation, kNN ignores one neighbour if more than 10% steps in its searching steps or prediction steps contain incident data. The anomaly-awareness evaluation is done in R programming language version 3.4.3 and the results are analyzed in Python programming language version 3.5.2. TSoutliers is in R library *forecast* version 8.2 while SH-ESD is in *AnomalyDetection* version 1.0.

C. Evaluation and Measurement

Stratified ten-fold cross-validation (CV) is used for the evaluation of anomaly-awareness predictions according to detected conditions, i.e. anomaly vs. normality. To have fair results, the final measurement and analysis are done considering the ground truth incident labels, i.e. incident vs. non-incident. As the evaluation requires lables, only one dataset with enough ground truth incident labels is used while we did not find other large enough datasets with all labels or sufficient information for labelling.

Mean absolute error (MAE) is used to measure the performance of predictions. MAE is selected as it is frequently used

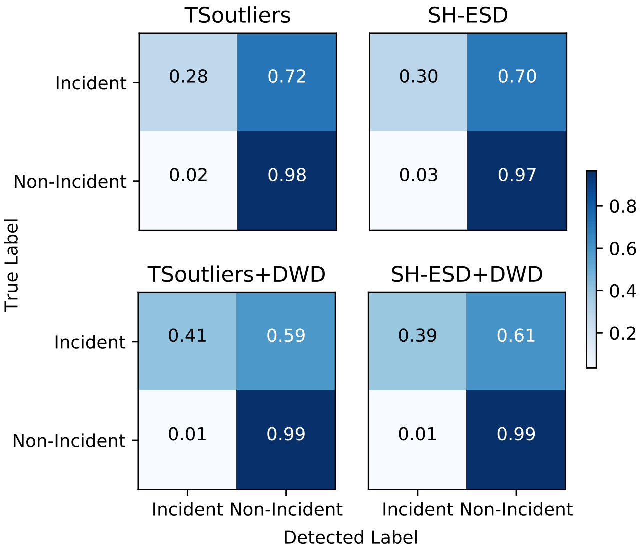


Fig. 3. Normalized confusion matrices of anomaly detection methods. The anomaly detection results are evaluated using incident labels. All true positive and true negtive ratios are increased while all false alarms are decreased.

in previous literature and it is easier for readers to compare with previous work. For instance, to measure the flow rate prediction results:

$$\text{MAE} = \frac{\sum_{\delta=1}^{q} |\hat{r}_\delta - r_\delta|}{q} \tag{2}$$

where $\hat{r}$ is the predicted flow rate, r is the true flow rate, and q is the number of records. Predictions for missing values cannot be evaluated, and are excluded from evaluation.

V. Results and Analysis

A. Anomaly Detection Methods

The ground truth incident labels are used to evaluate anomaly detection performance. As shown in Table I, the DWD preprocessed methods are better on all aspects than the ones without DWD. DWD methods increase detection ratio (a.k.a true positive ratio, TPR) by 31% - 44%, and reduced false alarm ratio (a.k.a false positive ratio, FPR) by 69% - 74%. DWD methods improve F1-score by 84.3% compared to the other two methods. The corresponding normalized confusion matrices are shown in Fig. 3.

TABLE I
Detection performance measured by true positive ratio, false positive ratio, true negative ratio, false negative ratio, F1-score.

Method	TPR	FPR	TNR	FNR	F1
TSoutliers	28.235	1.973	98.027	71.765	0.281
SH-ESD	29.916	3.446	96.554	70.084	0.233
TSoutliers+DWD	40.538	0.522	99.478	59.462	0.507
SH-ESD+DWD	39.193	1.083	98.917	60.807	0.438

B. Anomaly-Aware vs. Unaware Predictions

The prediction results are shown in Table II, there is no obvious difference among unaware results of different anomaly-unaware predictions. Different anomaly-unaware predictions are using different data divisions for 10-fold CV. They are similar and the maximum difference is no more than 0.48% compared to the ground truth. Thus, the CV results are robust.

The results patterns are easier to see in Fig. 4 and Fig. 5. Fig. 4 shows the prediction errors during incident traffic flow, while non-incident results are shown in Fig. 5. There are significant differences among aware methods during incident flow, but not during non-incident normal flow.

The relative changes of MAE are shown in Fig. 6 compared to ground truth unaware results. For non-incident data, the differences are ignorable, as the relative changes of MAE are minor considering the fluctuations of CV results (0.48%). For incident data prediction, the ground truth labels give the best results, up to 15.3% improvement, which is expected. The detection methods with DWD preprocessing preform well and reducing MAE by 4.87% on average and the maximum improvement is 8.96% for $m = 1$ by SH-ESD. The methods without DWD are giving less improvement (SH-ESD reduces MAE by only 2.36% during incident flow), or even increasing the prediction errors (TSoutliers increases MAE by 1.4%-6.9% during incident flow).

C. Analysis

One problem with the original methods (TSoutliers and SH-ESD) is that they treat high volume traffic flows during noon and holidays as abnormalities. It is due to the fact that they are not capturing daily and weekly seasonality correctly. Another problem is the high fluctuations and deviations during high traffic flow. DWD solves this problem by normalizing the residuals for different flow levels.

The results also motivate the improvement of anomaly detection algorithms. Improving the detection to be similar to the ground truth incident labels leads to incident-aware method and performs better. Currently, the detection ratio is not very high. While checking the results, two main reasons are found. First, some condition transition points as well as some points within incidents are similar to normal traffic flow volume and are not detected. Second, some minor incidents or the incidents that happened in other lanes are not detected.

The performance of anomaly detection algorithms influences the awareness prediction results significantly. The reason is that, removing information of time points that are from different traffic conditions leads to better fused results. For example, removing non-incident flow information when conducting incident flow prediction is actually removing 97.5% information, as only 2.5% data are incident data, according to the ground truth incident labels. However, removing incident flow information when conducting non-incident flow prediction is not giving significant improvements as the removed information is only 2.5%. Besides, the anomaly-aware methods require the underlying detection methods to perform well. If the detection algorithms are giving low accurate results, such

TABLE II
MAE OF PREDICTIONS. THE UNAWARE METHODS ARE SIMILAR AND CV IS ROBUST. FOR INCIDENT DATA, THE ANOMALY-AWARENESS AND DWD PREPROCESSING ARE USEFUL.

Awareness	Flow Condition	Detection Method	*m*=1	2	4	8
Aware	Incident	Ground Truth	12.662	13.944	15.822	17.548
		TSoutliers	15.983	16.476	17.281	18.258
		SH-ESD	14.598	15.329	16.626	17.950
		TSoutliers + DWD	13.705	14.695	16.205	17.766
		SH-ESD + DWD	13.611	14.575	16.172	17.723
	Non-Incident	Ground Truth	5.607	5.641	5.719	5.900
		TSoutliers	5.625	5.644	5.711	5.872
		SH-ESD	5.607	5.627	5.701	5.868
		TSoutliers + DWD	5.607	5.629	5.704	5.879
		SH-ESD + DWD	5.596	5.618	5.700	5.877
Unaware	Incident	Ground Truth	14.950	15.636	16.690	18.002
		TSoutliers	14.920	15.624	16.660	17.930
		SH-ESD	14.879	15.592	16.641	17.968
		TSoutliers + DWD	14.943	15.603	16.675	17.960
		SH-ESD + DWD	14.910	15.569	16.613	17.950
	Non-Incident	Ground Truth	5.607	5.632	5.702	5.878
		TSoutliers	5.606	5.630	5.701	5.877
		SH-ESD	5.606	5.630	5.703	5.880
		TSoutliers + DWD	5.607	5.629	5.704	5.879
		SH-ESD + DWD	5.607	5.629	5.700	5.873

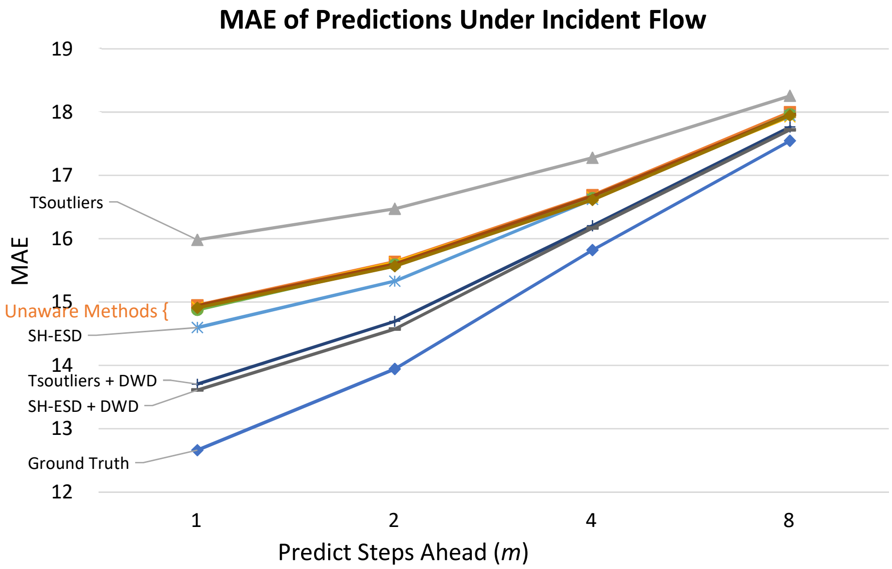


Fig. 4. Mean absolute error (MAE) of incident flow prediction. The incident-aware methods are distinguishable. The unaware methods are clustered together and not distinguishable which indicate cross-validated results are robust. The improvement of incident-aware methods with DWD are significant.

as the ones without DWD, the prediction accuracy will be also low, even worse than the anomaly-unaware methods. This is due to information is fused into wrong conditions.

VI. CONCLUSION AND DISCUSSION

The experiments results show that the proposed method of fusing information for traffic conditions separately to build different ensemble models can improve the prediction accuracy. Especially for incident conditions, the predictions are improved by 15.3%. The results are consistent with previous literature that different conditions should be treated separately for prediction. It also proves that our idea works well for fusing information in ensemble methods.

If no true labels are available to separate data conditions, anomaly detection is important. Our DWD enhanced methods act as alternatives to true labels and the predictions are improved by 8.96%. Although both TSoutliers and SH-ESD are supposed to be able to handle seasonal components, experiments show that our new DWD preprocessing provides better results. It increased detection ratio (TPR) by 31% - 44%, and reduced false alarm ratio (FPR) by 69% - 74%. The improvement comes from two reasons. First, the STL decomposition only removes one seasonality, while traffic data contains at least two main seasonality, daily and weekly. Second, the fluctuations and deviations on different flow levels are different, which are not considered in TSoutliers and SH-

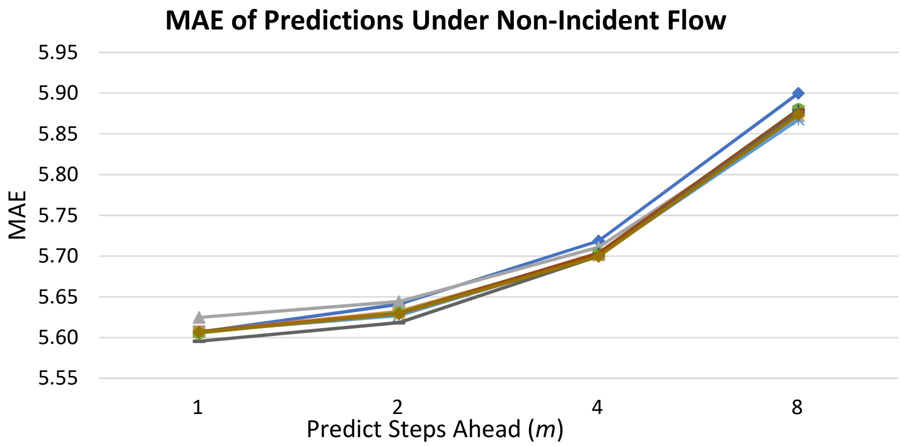


Fig. 5. MAE of non-incident flow prediction, both incident-aware and unaware. The differences among different methods are minor.

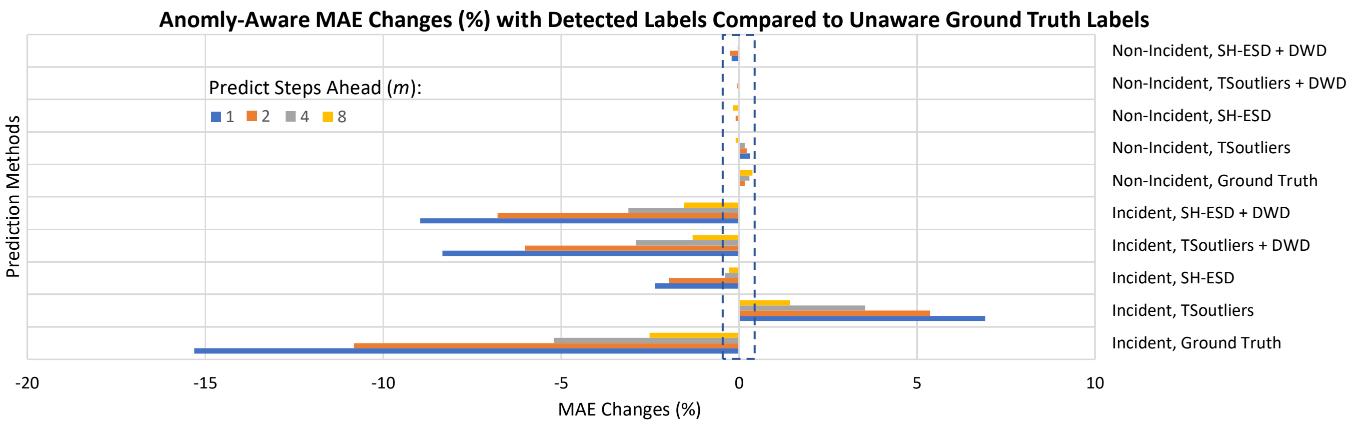


Fig. 6. Relative changes of anomaly-aware prediction MAE compared to ground truth unaware results. The changes during incident flow are significant, but not significant during non-incident flow. The values within the dash-circle is under the fluctuations of CV results (0.48%) and can be ignored. For incident flow, DWD based methods perform well and produce results similar to the ground truth. Results from other methods or conditions are not giving significant improvements, some even give worse results (higher MAE).

ESD.

Automation for anomaly detection is important due to the increasing amount of data. We also tried the classical SARIMA method. However, the automated SARIMA method from [29] cannot estimate the correct parameters. Another alternative one [28] find "no suitable ARIMA model", which also indicate the problem of SARIMA parameter choices.

Prediction of coming flow conditions (incident vs. non-incident) is topic that covers a wide variety of alternative methods.We plan to investigate and improve the automated ones in the future work.

Acknowledgment

We would like to thank reviewers for their valuable suggestions. This work is supported by National Natural Science Foundation of China (NSFC) No. 61364019 and Shandong Provincial Natural Science Foundation No. ZR2018PF009.

References

[1] A. M. de Souza, R. S. Yokoyama, G. Maia, A. Loureiro, and L. Villas, "Real-time path planning to prevent traffic jam through an intelligent transportation system," in *2016 IEEE Symposium on Computers and Communication (ISCC)*, Jun. 2016, pp. 726–731.

[2] S. An, T. Zhang, X. Zhang, and J. Wang, "Unrecorded Accidents Detection on Highways Based on Temporal Data Mining," *Mathematical Problems in Engineering*, vol. 2014, no. 852495, pp. 1–7, 2014.

[3] J. Guo, W. Huang, and B. M. Williams, "Real time traffic flow outlier detection using short-term traffic conditional variance prediction," *Transportation Research Part C-Emerging Technologies*, vol. 50, pp. 160–172, Jan. 2015.

[4] S. Oh, Y.-J. Byon, K. Jang, and H. Yeo, "Short-term Travel-time Prediction on Highway: A Review of the Data-driven Approach," *Transport Reviews*, vol. 35, no. 1, pp. 4–32, Jan. 2015.

[5] E. I. Vlahogianni, M. G. Karlaftis, and A. Stathopoulos, "An extreme value based neural clustering approach for identifying traffic states," in *IEEE Intelligent Transportation Systems Conference (ITSC)*. New York: IEEE, SEP 13-16, 2005.

[6] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Temporal evolution of short-term urban traffic flow: A nonlinear dynamics approach," *Computer-Aided Civil and Infrastructure Engineering*, vol. 23, no. 7, pp. 536–548, Oct. 2008.

[7] T.-I. Theodorou, A. Salamanis, D. D. Kehagias, D. Tzovaras, and C. Tjortjis, “Short-Term Traffic Prediction under Both Typical and Atypical Traffic Conditions using a Pattern Transition Model,” in *VEHITS*, 2017.

[8] R. Garcia-Rodenas, M. L. López-García, and M. T. Sánchez-Rico, “An Approach to Dynamical Classification of Daily Traffic Patterns: An approach to dynamical classification of daily traffic patterns,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 3, pp. 191–212, Mar. 2017.

[9] B. Sun, W. Cheng, G. Bai, and P. Goswami, “Correcting and Complementing Freeway Traffic Accident Data Using Mahalanobis Distance Based Outlier Detection,” *Tehnicki Vjesnik-Technical Gazette*, vol. 24, no. 5, pp. 1597–1607, Oct. 2017.

[10] L. Ma, S. Destercke, and Y. Wang, “Online active learning of decision trees with evidential data,” *Pattern Recognition*, vol. 52, no. Supplement C, pp. 33–45, Apr. 2016.

[11] W. Palma, “Missing Values and Outliers,” in *Time Series Analysis*. John Wiley & Sons, Apr. 2016, p. 424.

[12] S. I. Khan and S. G. Ritchie, “Statistical and neural classifiers to detect traffic operational problems on urban arterials,” *Transportation Research Part C-Emerging Technologies*, vol. 6, no. 5-6, pp. 291–314, OCT-DEC 1998.

[13] Y. Gu, Z. Qian, and F. Chen, “From Twitter to detector: Real-time traffic incident detection using social media data,” *Transportation Research Part C-Emerging Technologies*, vol. 67, pp. 321–342, Jun. 2016.

[14] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, “A review of novelty detection,” *Signal Processing*, vol. 99, pp. 215–249, Jun. 2014.

[15] B. Sun, W. Cheng, P. Goswami, and G. Bai, “Flow-Aware WPT k-Nearest Neighbours Regression for Short-Term Traffic Prediction,” in *IEEE Symposium on Computers and Communication (ISCC)*. Heraklion, Greece: IEEE, Jul. 2017, pp. 48–53.

[16] B. Sun, W. Cheng, P. Goswami, and G. Bai, “An Overview of Parameter and Data Strategies for k-Nearest Neighbours Based Short-Term Traffic Prediction,” in *ACM International Conference Proceeding Series ICITT/ICSET 2017*. ACM, Sep.-Oct. 2017, vol. 133326, pp. 68–74.

[17] L. Ma, B. Sun, and L. Ziyi, “Bagging Likelihood-Based Belief Decision Trees,” in *20th International Conference on Information Fusion (FUSION)*, Xi-An, China, Jul. 2017, pp. 1–6.

[18] L. Ma, B. Sun, and C. Han, “Learning Decision Forest from Evidential Data: The Random Training Set Sampling Approach,” in *4th International Conference on Systems and Informatics (ICSAI)*, Hangzhou, China, Nov. 2017.

[19] A. Ihler, J. Hutchins, and P. Smyth, “Adaptive Event Detection with Time-varying Poisson Processes,” in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’06. New York, NY, USA: ACM, 2006, pp. 207–216.

[20] M. Gupta, J. Gao, C. Aggarwal, and J. Han, “Outlier Detection for Temporal Data: A Survey,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 9, pp. 2250–2267, Sep. 2014.

[21] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, “LOF: Identifying Density-based Local Outliers,” in *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD ’00. New York, NY, USA: ACM, 2000, pp. 93–104.

[22] Z. Li, Z. Li, N. Yu, and S. Wen, “Locality-Based Visual Outlier Detection Algorithm for Time Series,” *Security and Communication Networks*, p. UNSP 1869787, 2017.

[23] J. Ma and S. Perkins, “Time-series novelty detection using one-class support vector machines,” in *Proceedings of the International Joint Conference on Neural Networks, 2003.*, vol. 3, Jul. 2003, pp. 1741–1745 vol.3.

[24] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation-Based Anomaly Detection,” *ACM Transactions on Knowledge Discovery from Data*, vol. 6, no. 1, p. 3, Mar. 2012.

[25] H. Cho, Y.-j. Kim, H. J. Jung, S.-W. Lee, and J. W. Lee, “OutlierD: An R package for outlier detection using quantile regression on mass spectrometry data,” *Bioinformatics*, vol. 24, no. 6, pp. 882–884, Mar. 2008.

[26] C. Chen and L.-M. Liu, “Joint Estimation of Model Parameters and Outlier Effects in Time Series,” *Journal of the American Statistical Association*, vol. 88, no. 421, pp. 284–297, Mar. 1993.

[27] M. Bernas, B. Placzek, P. Porwik, and T. Pamula, “Segmentation of vehicle detector data for improved k-nearest neighbours-based traffic flow prediction,” *IET Intelligent Transport Systems*, vol. 9, no. 3, pp. 264–274, Apr. 2015.

[28] J. Lopez-de Lacalle, “Tsoutliers - R Package for Detection of Outliers in time Series,” May 2017.

[29] R. J. Hyndman and Y. Khandakar, “Automatic time series forecasting: The forecast package for R,” *Journal of Statistical Software*, vol. 27, no. 3, pp. 1–22, Jul. 2008.

[30] R. J. Hyndman and G. Athanasopou los, “STL decomposition,” in *Forecasting: Principles and Practice*, 1st ed. OTexts, Oct. 2013.

[31] J. Hochenbaum, O. S. Vallis, and A. Kejariwal, “Automatic Anomaly Detection in the Cloud Via Statistical Learning,” *arXiv:1704.07706 [cs]*, Apr. 2017.

[32] B. Rosner, “Percentage Points for a Generalized ESD Many-Outlier Procedure,” *Technometrics*, vol. 25, no. 2, pp. 165–172, May 1983.

[33] S. Kelly and K. Ahmad, “Propagating Disaster Warnings on Social and Digital Media,” in *Intelligent Data Engineering and Automated Learning – IDEAL 2015*. Springer, Cham, Oct. 2015, pp. 475–484.

[34] L. Bodrog, M. Kajo, S. Kocsis, and B. Schultz, “A robust algorithm for anomaly detection in mobile networks,” in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sep. 2016, pp. 1–6.

[35] K. Jensen, T. V. Do, H. T. Nguyen, and A. Arnes, “Better Protection of SS7 Networks with Machine Learning,” in *2016 6th International Conference on IT Convergence and Security (ICITCS)*, Sep. 2016, pp. 1–7.

[36] TRB, *Highway Capacity Manual (HCM)*. Washington, DC: Transportation Research Board, 2010.

[37] A. Ruckstuhl, “Robust Fitting of Nonlinear Regression Models (nlrob) in Package robustbase: Basic Robust Statistics,” Nov. 2017.

[38] B. Sun, L. Ma, W. Cheng, W. Wen, P. Goswami, and G. Bai, “An Improved k-Nearest Neighbours Method for Traffic Time Series Imputation,” in *Chinese Automation Congress (CAC)*. Jinan, China: IEEE, Oct. 2017.

[39] B. Sun, W. Cheng, P. Goswami, and G. Bai, “Short-Term Traffic Forecasting Using Self-Adjusting k-Nearest Neighbours,” *IET Intelligent Transport Systems*, vol. 12, no. 1, pp. 41–48, Feb. 2018.

[40] Nvidia, *CUDA Programming Guide 8.0*. Nvidia, Sep. 2016.