City Traffic Prediction based on Real-time Traffic Information for Intelligent Transport Systems

Zilu Liang[#1], Yasushi Wakahara[#*2]
[#]Graduate School of Engineering, The University of Tokyo
[*]Information Technology Center, The University of Tokyo
Tokyo, Japan
[1]z.liang@cnl.t.u-tokyo.ac.jp
[2]wakahara@nc.u-tokyo.ac.jp

***Abstract*— Intelligent Transportation Systems (ITS) have been considered important technologies to mitigate urban traffic congestion. Accurate traffic prediction is one of the critical steps in the operation of an ITS. While techniques for traffic prediction have existed for many years, the research effort has mainly been focused on highway networks. Due to the fundamental difference between the traffic flow pattern on highways and that on city roads, much of the existing models cannot be effectively applied to city traffic prediction. In this paper, we propose two city traffic prediction models using different modeling approaches. Model-1 is based on the traffic flow propagation in the network, while Model-2 is based on the time-varied spare flow capacity on the concerned road link. The proposed models are implemented to predict the traffic volume in Cologne in Germany, and the real data are collected through simulations in the traffic simulator SUMO. The results show that both of the proposed models reduce the prediction error up to 52% and 30% in the best cases compared to the existing Shift Model. In addition, we found that Model-1 is suitable for short prediction interval that is in the same magnitude as the link travel time, while Model-2 demonstrates superiority when the prediction interval is larger than one minute.**

Keywords-urban traffic prediction, traffic volume, spatio-temporal correlation, intelligent transport system

I. INTRODUCTION

With more and more vehicles running on the roads, traffic congestion has continuously been a serious social problem faced by modern people. Being the combination of information communication technology and road traffic engineering, the Intelligent Transportation System (ITS) technology has caught world-wide attention because of its great potential in mitigating road traffic congestion at low cost. The traffic information provided by ITS can reflect either prevailing network conditions or predictive network states. It has been widely recognized in the transportation science community [1] that when predictions are accurate, predictive information is generally expected to be more effective than prevailing information because it accounts for the rapid change of traffic conditions spatially and temporally. For example, the optimization of traffic signal control or variable message signs should be based on expected traffic conditions in the near future rather than based on a traffic situation soon to be obsolete [2]. Similarly, a traveler would prefer to receive route guidance information corresponding to the likely traffic when he/she will be on the road rather than that occurred prior to the start of his/her trip [3]. Therefore, traffic prediction is a critical step in the efficient operation of ITS. Intensive research effort has been made on traffic prediction on highways, but the existing work on traffic prediction on city roads is limited. The features of a city traffic network include: 1) the road links are short and the number of links is large, thus the prediction on traffic situation on all links is potentially computationally costly; 2) traffic flow is frequently split up due to the prevalent existence of intersections; 3) the traffic management at intersections, e.g. traffic signal control, has strong impact on traffic flow pattern, and the occurrence of traffic congestion is more accurately indicated by traffic volume on road links rather than travel speed. Due to the above intrinsic difference between a city traffic network and a highway system, most of the existing highway traffic prediction models cannot be effectively applied to city roads.

In this paper, we seek to predict traffic volume on city roads by adopting two distinct modeling approaches. The first model is based on the propagation of traffic flow along the successive road links on a route, while the second model is based on time-varied spare flow capacity on the concerned links. The validity of the models is ensured by incorporating real-time traffic information which reflects the up-to-date traffic situation in the network. Hence, the proposed models are adaptive to the dynamics change of traffic flow pattern. In addition, the proposed models have low requirement on historical data, and they are easy to implement and are suitable to be embedded to traffic management strategies including traffic signal control and in-vehicle route guidance. The performance of the proposed models is evaluated through the prediction of traffic volume in Cologne, Germany, and the real traffic data are acquired by traffic simulations in the microscopic traffic simulator SUMO [21]. The influence of prediction interval over the prediction error is studied to investigate the characteristics of each model.

978-1-4799-0846-2/13/$31.00 ©2013 IEEE

978-1-4799-0846-2/13/$31.00 ©2013 IEEE

The rest of the paper is organized as follows. In Section 2 we discuss related works. In Section 3 we present the proposed city traffic prediction models. In Section 4, we demonstrate the evaluation of the proposed models based on the prediction results and discuss the characteristics of the models. In the last section we draw conclusions.

II. RELATED WORKS

Intensive research effort has been made on traffic prediction so far, mostly in highway scenarios. Generally speaking, the existing works on traffic prediction can be classified into two categories: parametric methods and non-parametric methods. Parametric methods include prediction models based on historical average and smoothing techniques, regression [4], Kalman filter [5], autoregressive integrated moving average (ARIMA) [6][7], etc. Nonparametric methods refer to the prediction models using non-parametric regression [8][9], and artificial intelligence techniques such as machine learning [10], fuzzy logic [11][12] and neural networks [13][14]. Smith and Demetsky [15] conducted comparisons of historical average, time-series, nonparametric regression, and ANN and found that the non-parametric regression model significantly outperformed the other models and was easier to implement. Even so, non-parametric regression models require large amount of historical data, and it requires training processes. Moreover, in the scenario where not enough good matches are found in the historical database, the non-parametric regression may fail to output a reliable prediction. Therefore, there is a need to develop new traffic prediction model for city traffic prediction, and the model is expect to be 1) easy to implement; 2) computationally friendly and able to perform network-wide prediction in a real-time manner; 3) weakly dependent on historical traffic data; 4) is adaptive to the dynamics of traffic situation. As such, the goal of this paper is to propose effective city traffic prediction models that satisfy the above requirements.

III. THE PROPOSED PREDICTION MODELS

In the proposed city traffic prediction model, the city traffic system is considered as a discrete-time dynamic system. The rolling horizon approach [16] is adopted to conduct the prediction. In other words, the time horizon is divided into discrete traffic prediction time intervals whose length is τ seconds, and traffic prediction is performed recursively every τ seconds and at the beginning of each time interval. The notations that are used in the prediction model are summarized in TABLE I.

Before presenting the details of the city traffic prediction models, we first clarify the assumptions for the models here. We assume that the vehicles are uniformly distributed on road links and are traveling at constant speed; that is, acceleration and deceleration of vehicles are not considered. The traffic volume is assumed to stay constant during a prediction time interval. Moreover, we assume that the split rate of traffic flows at the intersections are obtained beforehand e.g. via vehicle tracking [17].

TABLE I. NOTATIONS USED IN THE PREDICTION MODEL

Variable	Description
i	Road link
L_i	The length of link i (m)
C_i	The capacity of link i (veh)
$\bar{v}_i$	The average speed on link i in time interval t (m/s)
U_i	The number of upstream links on link i
D_i	The number of downstream links on link i
$(i-1)_{u_i}$	The u-th upstream link of link i, $u_i = 1,2,\dots,U_i$
$(i+1)_{d_i}$	The d-th downstream link of link i, $d_i = 1,2,\dots,D_i$
τ	The prediction interval (s)
L_v	The average length of vehicles (m)
L_g	The minimum gap between vehicles (m)
$X_i(t)$	The real traffic volume on link i in time interval t (veh)
$\hat{X}_i(t)$	The predicted traffic volume on link i in time interval t (veh)
$\bar{X}_i(t)$	The average traffic volume on link i in time interval t (veh)
$\hat{X}_{dep,i}(t)$	The traffic volume that starts the trip on link i in time interval t (veh)
$\hat{X}_{arr,i}(t)$	The traffic volume that ends the trip on link i in time interval t (veh)
$\hat{Q}_{i,in}(t)$	The predicted traffic volume that enters link i during time interval t (veh)
$\hat{Q}_{i,out}(t)$	The predicted traffic volume that enters link i during time interval t (veh)
$S_{i\to(i+1)_d,out}(t)$	The saturated outflow of a road link i to its d-th downstream link $i+1$ in time interval t (veh)
$S_{i,in}(t)$	The saturated inflow of link i in time interval t (veh)
$\gamma_{(i-1)_u\to i}(t)$	The split rate of traffic volume that travels on the u-th upstream of link i in time interval t and will enter link i afterwards (%)
$\gamma_{i\to(i+1)_d}(t)$	The split rate of traffic volume that travels on the d-th downstream link in time interval t and will enter link i afterwards (%)
$\delta_{(i-1)_u\to i}(t)$	The adjustment factor for the traffic flow from the u-th upstream link to link i in time interval t
$\delta_{i\to(i+1)_d}(t)$	The adjustment factor for the traffic flow from link i to its d-th downstream link in time interval t
$T^g_{(i-1)_u\to i}$	The ratio of green phase of the traffic signal for the traffic flow from the u-th upstream link to link i in time interval t (s)
$T^g_{i\to(i+1)_d}$	The ratio of green phase of the traffic signal for the traffic flow from link i to its d-th downstream link in time interval t (s)

A. Model-1: Prediction based on Traffic Flow Propagation

The Model-1 is based on the rationale that traffic flow propagates from upstream road links to downstream road links in a traffic network. In addition, the traffic flow may be split up at intersections depending on the predetermined destination of each vehicle and the route selected by the drivers.

$$\hat{X}_i(t+1) = \max\{X_i(t) + \hat{Q}_{i,in}(t) + X_{dep,i}(t) - \hat{Q}_{i,out}(t) - X_{arr,i}(t), 0\}, \quad (1)$$

where

$$\hat{Q}_{i,in}(t) = \sum_{u=1}^{U_i} X_{(i-1)_u}(t) \cdot \gamma_{(i-1)_u \to i}(t) \cdot \delta_{(i-1)_u \to i}(t) \quad (2)$$

$$\hat{Q}_{i,out}(t) = \sum_{d=1}^{D_i} X_i(t) \cdot \gamma_{i \to (i+1)_d}(t) \cdot \delta_{i \to (i+1)_d}(t) \quad (3)$$

$$\delta_{(i-1)_u \to i}(t) = \min\{\frac{T^g_{(i-1)_u \to i}(t)}{\widehat{TT}_{(i-1)_u}(t)}, 1\} \quad (4)$$

$$\delta_{i \to (i+1)_d}(t) = \min\{\frac{T^g_{i \to (i+1)_d}(t)}{\widehat{TT}_i(t)}, 1\} \quad (5)$$

For simplicity, we estimate the link travel time using Equation (6) in the implementation of this model, though we are aware of the possibility to improve the accuracy by adopting more comprehensive travel time estimation algorithms.

$$\widehat{TT}_i(t) = \frac{L_i}{\bar{v}_i(t)} \quad (6)$$

B. Model-2: Prediction based on Spare Road Capacity

The Model-2 is based on the rationale that the inflow and outflow rate on the concerned link is largely determined by its spare capacity. In order to quantify the time-varied spare capacity of a road link, we first define the saturated outflow and inflow of a link. The saturated outflow of a road link i to its d-th downstream link $(i+1)_d$, denoted by $S_{i\to(i+1)_d,out}(t)$, is defined as the maximum number of vehicles per hour that can exit link i and enter link $i+1$. Similarly, the saturated inflow of link i, denoted by $S_{i,in}(t)$, is defined as the maximum number of vehicles per hour that can enter link i from all its upstream links. The saturated outflow and inflow are determined by the real-time traffic situation on the concerned link, which may vary significantly at different times. The saturate outflow $S_{i\to(i+1)_d,out}(t)$ and saturated inflow $S_{i,in}(t)$ can be approximately calculated by the following formulas:

$$S_{i\to(i+1)_d,out}(t) = \frac{T^g_{i\to(i+1)_d}\bar{v}_i(t)}{L_v + L_g} \quad (7)$$

$$S_{i,in}(t) = \frac{\bar{v}_i(t)}{L_v + L_g} \quad (8)$$

Accordingly, the spare capacity of a link in terms of inflow and outflow is an adjustment of the saturated flows based on the real-time vehicle occupancy on the concerned links. Based on the above consideration, the Model-2 is formulated below.

$$\hat{X}_i(t+1) = \max\{X_i(t) + \hat{Q}_{i,in}(t) + X_{dep,i}(t) - \hat{Q}_{i,out}(t) - X_{arr,i}(t), 0\}, \quad (9)$$

where

$$\hat{Q}_{i,in}(t) = \min\{\sum_{u=1}^{U_i} S_{(i-1)_u \to i,out}(t) \cdot \frac{\bar{X}_{(i-1)_u}(t)}{C_{(i-1)_u}} \cdot \gamma_{(i-1)_u \to i}(t), S_{i,in}(t)\} \quad (10)$$

$$\hat{Q}_{i,out}(t) = \sum_{d=1}^{D_i} \min\{S_{i\to(i+1)_d,out}(t) \cdot \frac{\bar{X}_i(t)}{C_i} \cdot \gamma_{i\to(i+1)_d}(t), S_{(i+1)_d,in}(t)\} \quad (11)$$

The saturated flows are calculated according to the definitions by Equation (7) and (8).

$$S_{(i-1)_u\to i,out}(t) = \frac{T^g_{(i-1)_u\to i}(t) \cdot \bar{v}_{(i-1)_u}(t)}{\tau \cdot (L_v + L_g)} \quad (12)$$

$$S_{i\to(i+1)_d,out}(t) = \frac{T^g_{i\to(i+1)_d}(t) \cdot \bar{v}_i(t)}{\tau \cdot (L_v + L_g)} \quad (13)$$

$$S_{i,in}(t) = \frac{\bar{v}_i(t)}{\tau \cdot (L_v + L_g)} \quad (14)$$

$$S_{(i+1)_d,in}(t) = \frac{\bar{v}_{(i+1)_d}(t)}{\tau \cdot (L_v + L_g)} \quad (15)$$

IV. Performance Evaluation

In order to evaluate the performance of the proposed city traffic prediction model, we implement the models to predict traffic volume in the Cologne, Germany. The microscopic traffic simulator SUMO [19] is employed to simulate and collect the real traffic volume on individual road links, mainly because of the unavailability of the real traffic volume data in city scenarios. We input the real data of traffic demand between 6:00 and 8:00 a.m. of a day in the "TAPAS Cologne" Scenario [18] to the simulator and collect the traffic volume on the concerned links as the real values for our evaluation. The default setting in SUMO 0.15.0 is used to configure vehicles. The vehicle length is

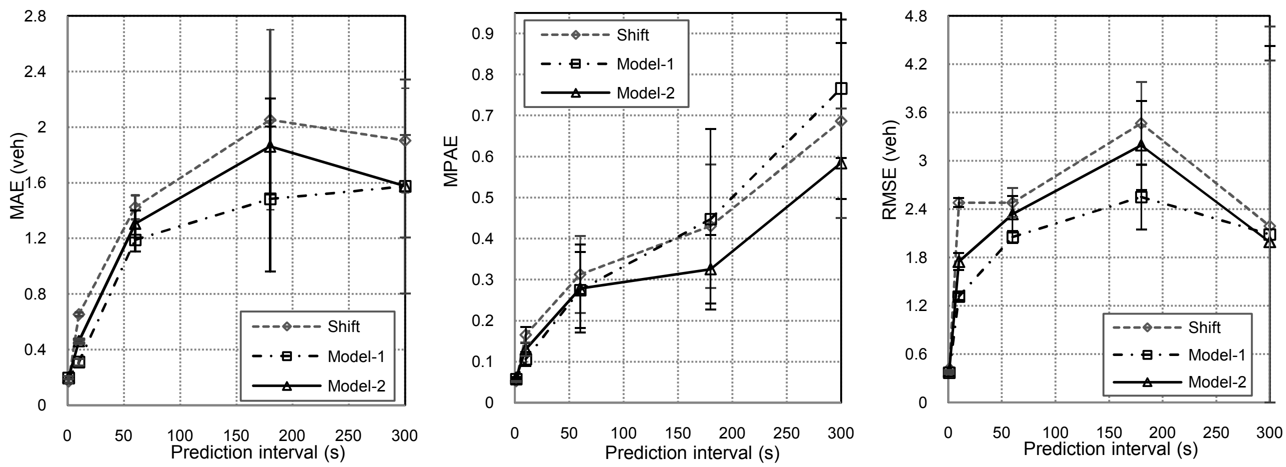

Figure 1. Error analysis. (left) Mean Absolute Error (MAE); (center) Mean Absolute Percent Error (MAPE); (right) Root Mean Square Error (RMSE).

5m, the minimal gap between vehicles is 2.5m, and the Krauss model [20] is used as car following model. We intend to predict the traffic volume on link 23572355#2 which has two upstream links and three downstream links. The parameters of the local topology of link 23572355#2 are shown in TABLE II.

TABLE II. THE PARAMETERS OF THE LOCAL TOPOLOGY LINK23572355#2

Link ID	Length (m)	Max. Speed (m/s)	Relation
23572355#2	109	13.89	Link $\boldsymbol{i}$
-24665807#1	125	13.89	Upstream link of i
24665807#0	134	13.89	Upstream link of i
23572355#1	98	13.89	Downstream link of i
23585509#0	103	13.89	Downstream link of i
23572355#3	121	13.89	Downstream link of i

We compare our proposed models with the existing Shift Model, which is represented by Equation (16)

$$\hat{X}_i(t+1) \ = \ X_i(t) \tag{16}$$

Since the prediction interval is small, the Shift Model is considered a very competitive prediction model for dynamic traffic networks. We employ the following measures as criteria to evaluate the accuracy of the prediction models: Mean Absolute Error (MAE), Mean Absolute Percent Error (MAPE), and Root Mean Square Error (RMSE). The definitions of these measures are shown below.

$$MAE \ = \frac{1}{T} \sum_{t=1}^{T} |\hat{X}(t) - \ X(t)| \tag{17}$$

$$MAPE \ = \frac{1}{T} \sum_{t=1}^{T} \frac{|\hat{X}(t) - X(t)|}{X(t)} \times 100\% \tag{18}$$

$$RMSE \ = \ \sqrt{\frac{1}{T} \sum_{t=1}^{T} |\hat{X}(t) - \ X(t)|^2} \tag{19}$$

where T is the total number of prediction intervals, $X(t)$ are the values collected from the simulations in SUMO, while $\hat{X}(t)$ are the predicted values. We conduct prediction every 1, 10, 60, 180, and 300 seconds, which is equivalent to the aggregation period of traffic data. We run the simulations 5 times under different seed values, and acquire 5 sets of real traffic data. For each value of prediction interval, the prediction is conducted over each of the 5 sets of data and prediction errors are calculated after each prediction. The average values of the prediction errors over the 5 repetitions are taken as the final results, and the 95% confidence interval is also calculated.

Figure 1 demonstrates the effect of prediction interval on the prediction accuracy of each model. Generally speaking, the prediction errors increase as the prediction interval increases. Model-1 significantly reduces MAE by 52%, MAPE by 36% and RMSE by 47% compared to the baseline Shift Model in the best case (prediction interval is 10s). Model-2 also reduces MAE by 30% (prediction interval is 10s), MAPE by 24% (prediction interval is 180s) and RMSE by 30% (prediction interval is 10s) compared to the Shift Model in the best cases. With respect to MAPE, Model-1 gains its maximum advantage when the prediction interval is 10s, which is in the same magnitude as the link travel time. As the prediction interval further increases, the prediction accuracy of Model-1 is not satisfactory. In comparison, Model-2 consistently demonstrates superiority as the prediction interval is larger than one minute.

In practice, the aggregation period of traffic data ranges from 20 or 30 seconds [20] to 5 minutes. The prediction model should be chosen depending on the data aggregation period. For example, if the traffic data is aggregated less than 1 minute, Model-1 should be used to perform the prediction; otherwise, Model-2 would be a better candidate to yield accurate prediction. In addition, it is worthy of mentioning that the prediction accuracy could be influenced by other factors, such as the characteristics of traffic demand, the topology of the traffic network, the route choice decision made by drivers with or without guidance, the configuration of traffic signal, etc. Hence, the ultimate requirement on the prediction accuracy could be greatly dependent on the specific applications. Figure 2 and Figure 3 further indicate that the proposed models are not biased.

V. CONCLUSIONS

In this paper, we proposed two city traffic prediction models based on the propagation of traffic flow and the spare road capacity respectively. We evaluated the proposed models by comparing their performance with the Shift Model under varied prediction interval using the real data collected in the traffic simulator SUMO. The preliminary results demonstrated that both models significantly reduce prediction error up to 52% and 30% in the best cases compared to the existing Shift Model, and they are expected to produce good prediction accuracy in real situations. In addition, we found that the performance of Model-1 peaks when the prediction interval is in the same magnitude as the link travel time, while Model-2 demonstrates superiority when the prediction interval is larger than one minute. Depending on the aggregation period of traffic data, different model should be selected to generate the best prediction accuracy.

In the next step, we intend to further investigate the effects of other parameters on the performance of the proposed models, including the topology of the traffic network and the characteristics of the traffic demand. We will also study the possibility of constructing a hybrid model to combine the merits of both models. Being aware of the fact that the required prediction accuracy is closely related to the purpose of specific applications, we will also conduct case studies to investigate the relation between the prediction accuracy and the final objectives of the applications.

REFERENCES

[1] J. Dong, H. S. Mahmassani, C.-C. Lu, “How Reliable Is This Route?: Predictive Travel Time and Reliability for Anticipatory Traveler Information Systems,” Transportation Research Record, vol.1980, iss.1, pp.117-125, 2006.

[2] T. Kato, Y. Kim, T. Suzuki, and S. Okuma, “Model predictive control of traffic flow based on hybrid system modeling,” IEICE Trans. on Fundamentals, vol.E88-A, no.2, pp.549-560, 2005.

[3] K. Chen, S. E. Underwood, “Research on anticipatory route guidance,” Vehicle Navigation and Information Systems Conference, vol.2, pp.427-439, 1991.

[4] X. Zhang, and J. A. Rice, “Short-term travel time prediction,” Transportation Research Part C, vol.11, pp.187-210, 2003.

[5] Y.Wang, M.Papageorgiou, and A.Messmer, “Real-time freeway traffic state estimation based on extended Kalman filter: a case study,” Transportation Science, vol.41, no.2, p.167-181, 2007.

[6] W. Min, and L. Wynter, “Real-time road traffic prediction with spatio-temporal correlations,” Transportation Research Part C, vol.19, pp.606-616, 2011.

[7] J. Xia, M. Chen, and W. Huang, “A multistep corridor travel-time prediction method using presence-type vehicle detector data,” Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, vol.15, no.2, pp.104-113, 2011.

[8] H. Markovic, et.al, “GPS data based non-parametric regression for predicting travel time in urban traffic networks,” Science in Traffic and Transport, vol.22, no.1, pp.1-13, 2010.

[9] C. H. Wu, J. M. Ho, and D. T. Lee, “Travel-time prediction with support vector regression,” IEEE Transactions on Intelligent Transportation Systems, vol.5, no.4, pp.276-281, 2004.

[10] J. Rzeszotko and S. H.Nguyen, “Machine Learning for Traffic Prediction,” Fundam. Inf., vol.119, no.3-4, p.407-420, August 2012.

[11] L. Li, W.-H. Lin 1, H. Liu, “Type-2 fuzzy logic approach for short-term traffic forecasting,” IEE Proceedings - Intelligent Transport Systems, vol.153, no.1,pp.33-40, March 2006.

[12] V. P. Vijayan, B. Paul, “Multi Objective Traffic Prediction Using Type-2 Fuzzy Logic and Ambient Intelligence,” Advances in Computer Engineering (ACE), 2010 International Conference on , pp.309-311, June 2010.

[13] E. Mazloumi, G. Roase, G. Currie, et.al, “Prediction intervals to account for uncertainties in neural network predictions: Methodology and application in bus travel time prediction,” Engineering Applications of Artificial Intelligence, vol.24, no.3, pp.534-542, April 2011.

[14] C. van Hinsbergen, A. Hegyi, J. van Lint, and H. van, Zuylen, “Bayesian neural networks for the prediction of stochastic travel times in urban networks,” IET Intelligent Transport Systems, vol.5, no.4, pp.259-265, 2011.

[15] B.L.Smith, B.M.Williams, and R.K.Oswalsd, “Comparison of parametric and nonparametric models for traffic flow forecasting,” Transportation Research Part C, vol.10, no.4, p.303-321, 2002.

[16] T.L.Pan, A.Sumalee, R.X.Zhong, N.Indra-payoong, “Short-Term Traffic State Prediction Based on Temporal–Spatial Correlation," IEEE Transactions on *Intelligent Transportation Systems,* vol.PP, no.99, 2012.

[17] J. Hu, A. Razdan, J. Femiani, et.al, “Road network extraction and intersection detection from aerial images by tracking road footprints,” IEEE T. Geoscience and Remote Sensing, vol.45, no.12-2, p.4144-4157, 2007.

[18] “TAPAS Cologne” Scenario. http://sumo.sourceforge.net/doc/current/docs/userdoc/Da

[19] D. krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, ”Recent development and applications of SUMO – Simulation of Urban MObility,” International Journal on Advances in Systems and Measurements, vol.5, no.3&4, pp.128-138, 2012.

[20] M. L. Hazelton, “Estimating vehicle speed from traffic count and occupancy data,” Journal of Data Science, vol.2, p.231-244, 2004.

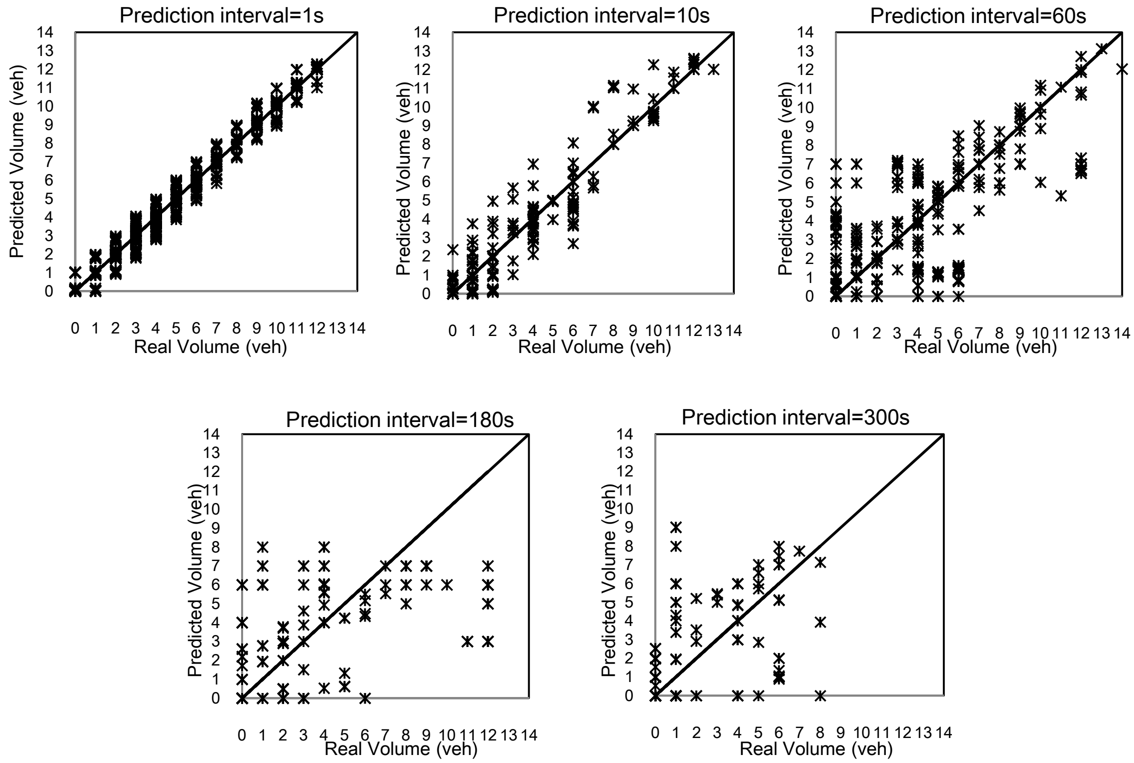

Figure 2. Real values v.s. predicted values of Model-1

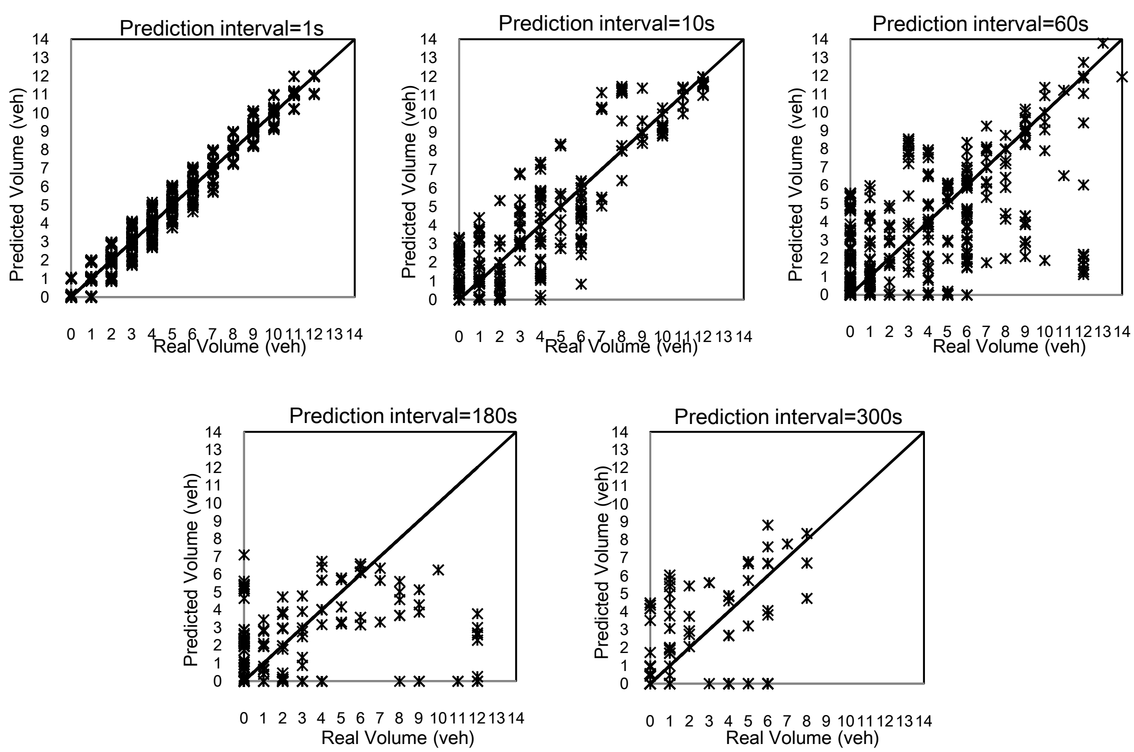

Figure 3. Real values v.s. predicted values of Model-2.