

Research Article

Three Revised Kalman Filtering Models for Short-Term Rail Transit Passenger Flow Prediction

Pengpeng Jiao,¹ Ruimin Li,² Tuo Sun,¹ Zenghao Hou,³ and Amir Ibrahim⁴

¹Beijing Urban Transportation Infrastructure Engineering Technology Research Center, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

²Institute of Transportation Engineering, Tsinghua University, Beijing 100084, China

³Parsons Transportation Group, 100 Broadway, New York, NY 10005, USA

⁴New Jersey Department of Transportation (NJDOT), 1035 Parkway Avenue, Trenton, NJ 08625, USA

Correspondence should be addressed to Pengpeng Jiao; jiaopengpeng@bucea.edu.cn

Received 16 December 2015; Accepted 10 March 2016

Academic Editor: Payman Jalali

Copyright © 2016 Pengpeng Jiao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Short-term prediction of passenger flow is very important for the operation and management of a rail transit system. Based on the traditional Kalman filtering method, this paper puts forward three revised models for real-time passenger flow forecasting. First, the paper introduces the historical prediction error into the measurement equation and formulates a revised Kalman filtering model based on error correction coefficient (KF-ECC). Second, this paper employs the deviation between real-time passenger flow and corresponding historical data as state variable and presents a revised Kalman filtering model based on Historical Deviation (KF-HD). Third, the paper integrates nonparametric regression forecast into the traditional Kalman filtering method using a Bayesian combined technique and puts forward a revised Kalman filtering model based on Bayesian combination and nonparametric regression (KF-BCNR). A case study is implemented using statistical passenger flow data of rail transit line 13 in Beijing during a one-month period. The reported prediction results show that KF-ECC improves the applicability to historical trend, KF-HD achieves excellent accuracy and stability, and KF-BCNR yields the best performances. Comparisons among different periods further indicate that results during peak periods outperform those during nonpeak periods. All three revised models are accurate and stable enough for on-line predictions, especially during the peak periods.

1. Introduction

With the rapid development of urbanization and motorization in most Chinese large cities, the urban transportation systems are facing more and more serious problems, such as congestion, crashes, and pollution. As an efficient trip mode, rail transit system has played a more and more important role in solving traffic issues. In Beijing, there are a total of 21 lines in operation now, covering a distance of 527.2 kilometers (327.6 miles). During the past decade, the average daily passenger flow has increased dramatically to about 10 million riders. Therefore, the operation and management of the rail transit system, especially real-time operation, is very important.

During peak hours, pedestrian congestion happens frequently. For safe and efficient purposes, the real-time

passenger flows, especially predicted flows during the next several time intervals, are key issues for real-time intelligent operation of the rail transit system. However, with the past and current passenger flows detected easily, the future flows are not straightforward. Therefore, the passenger flow forecast method based on statistical data is rather meritorious.

Most recently, Sun et al. [1] proposed a nonparametric regression method to forecast passenger flow at subway transfer stations. Except for this, the literature review shows that very few researches have focused directly on short-term rail transit passenger flow prediction. However, short-term traffic flow forecasting has been studied extensively with Intelligent Transportation Systems (ITS) and many practical models have been developed from these studies. With just different input data entered into these models, some of those

achievements can be used to forecast rail transit passenger flow easily.

Existing traffic flow forecast models cover a wide range consisting of Historical Average (HA), Autoregressive Integrated Moving Average (ARIMA), Neural Network (NN), Kalman filtering (KF), nonparametric regression (NR), chaos theory, Support Vector Machine (SVM), and others. The HA model uses a simple time-series method [2], which is rarely in use now. Ahmed and Cook [3] put forward an ARIMA model to forecast freeway traffic flows, and Williams et al. [4] further developed it to seasonal case and compared it with an Exponential Smoothing Method (ESM). Many researchers formulated NN-based prediction models and obtained rather satisfying results such as Smith and Demetsky [5], Florio and Mussone [6], Zhang et al. [7], Dougherty et al. [8], Park and Rilett [9], and Vlahogianni et al. [10]. Kalman filtering is a kind of recursive state forecast method with high efficiency that has also been widely used in short-term traffic flow prediction, for example, Okutani and Stephanedes [11], Cathey and Dailey [12], and Shekhar and Williams [13]. As a nonlinear regression method, the NR model is rather applicable to uncertain and dynamic systems, just like real-time transportation systems. Pioneering work on the NR method can be found in Yakowitz [14] and Karlsson and Yakowitz [15], and some scholars further developed them for traffic flow forecast, for instance, Davis and Nihan [16], Smith and Demetsky [17], Oswald et al. [18], Smith et al. [19], Qi and Smith [20], and Kindzerske and Ni [21]. Huang et al. [22], Lu and Wang [23], Meng and Peng [24], Xue and Shi [25], and Pang and Zhao [26] applied chaos theory in the traffic flow prediction and obtained acceptable results. SVM is a new statistical machine-learning method [27] which has been proved to have stronger learning and generalization abilities than the NN model. SVM has also been used in the field of traffic flow forecast such as Ren et al. [28], Wu et al. [29], and Wang et al. [30].

Generally, the above methods can be classified into statistical and artificial intelligence models. Smith and Demetsky [17] and Smith et al. [19] compared some of these models and proposed that no single method was universally accepted as the best one. Therefore, based on existing single models, some combined methods have been developed and one of the most effective approaches is the Bayesian combined model. Zheng et al. [31], Dong et al. [32], Jiao et al. [33], and Jiao et al. [34] have proved its effectiveness.

More recently, some researches proposed new models for multistep prediction [35] and large-scale road network forecast [36]. The latter employed cloud computing techniques for large-scale network applications.

Among all the above short-term traffic flow forecast models, the Kalman filtering method is very efficient due to its recursive attribute and is rather convenient for use in rail transit passenger flow predictions. However, existing researches have proved that the traditional KF methods are not accurate and stable enough for on-line applications. Therefore, this paper will revise the traditional KF methods and propose three revised models.

To predict passenger flow accurately and efficiently, one key feature of the paper is to introduce some error calibration

measures or new state variables into classical models and to construct some revised KF forecast models. The second key feature is to integrate some stable methods and formulate an innovative KF prediction model with good accuracy, stability, and robustness.

This paper consists of six sections. Following the Introduction, the basic KF model is described in the second section, including its state transition and measurement equations. Three revised KF models are formulated in the third section, including the KF model based on the error correction coefficient (KF-ECC), the KF model based on Historical Deviation (KF-HD), and the KF model based on the Bayesian combination and nonparametric regression (KF-BCNR). Solution algorithms for the NR model, KF model, and Bayesian combination model are designed in the fourth section, respectively. Prediction results using practical statistical passenger flow data are reported and analyzed in the fifth section. Conclusions and some future research directions are summarized in the last section.

2. Basic Kalman Filtering Model

The KF model is a kind of state space method consisting of three important parts: state variable, state transition equation, and measurement equation.

In the rail transit passenger flow prediction, the short-term passenger flow to be forecasted is taken as the state variable directly. In this paper, we employ the passenger flow at the station. Using $Q(k)$ to denote the passenger flow during time interval k at a station, the state transition equation and measurement equation are formulated as follows:

$$\mathbf{Q}(k) = \mathbf{Q}(k-1) + \mathbf{W}(k), \quad (1)$$

$$\mathbf{H}(k) = \mathbf{M}(k) * \mathbf{Q}(k) + \mathbf{e}(k), \quad (2)$$

where $\mathbf{Q}(k)$ is column vector form of passenger flow $Q(k)$ and, accordingly, $\mathbf{Q}(k-1)$ is the column vector of $Q(k-1)$; $\mathbf{W}(k)$ is Gauss white noise vector with mean value $\mathbf{0}$ and covariance matrix $\mathbf{D}\delta_{ij}$ and here \mathbf{D} is a constant semipositive matrix and δ_{ij} is the Kronecker delta; that is, $\delta_{ij} = \{1, i = j; 0, \text{otherwise}\}$; $\mathbf{H}(k)$ is column vector form of measurements and here the Historical Average passenger flow during the same time interval k is taken as the measurement; $\mathbf{M}(k)$ is measurement matrix and here it equals the identity matrix in the passenger flow prediction; that is, it can be neglected in the formulation; $\mathbf{e}(k)$ is column vector form of detection errors with mean value $\mathbf{0}$ and covariance matrix $\mathbf{R}\delta_{ij}$ and here \mathbf{R} is a constant semipositive matrix similar to \mathbf{D} .

Equations (1) and (2) constitute the basic KF model together. Existing researches have proved that the basic form of KF is rather efficient due to its recursive attribute. However, the accuracy is not satisfying. Therefore, we further formulate some revised KF models to improve the prediction accuracy.

3. Three Revised Kalman Filtering Models

3.1. The Revised KF Model Based on Error Correction Coefficient. Since the historical passenger flow data could be

collected easily, we can conveniently track the trend of the flow changes. The basic KF model in (1) and (2) has been employed in historical cases, and the errors between historical forecast and historical detection are thus obtained. Based on characteristics of such errors, we introduce an error correction coefficient into the measurement equation:

$$\mathbf{H}(k) = \lambda \mathbf{Q}(k) + \mathbf{e}(k), \quad (3)$$

where λ is the error correction coefficient based on historical forecasting deviations. Here, measurement matrix $\mathbf{M}(k)$ is neglected, because it is an identity matrix in nature.

The error correction coefficient λ varies under different conditions. It is closely correlated to the historical forecasting errors. In detail, it grows with the increase of historical errors, and we can obtain it by the historical data fitting procedures.

During weekdays, rail transit passenger flows usually change from morning peak hours to nonpeak hours and then to evening peak hours. Therefore, some similar characteristics in the historical forecasting errors are observed. Statistical analyses prove that it can fit a quadratic parabola function:

$$\lambda = bk - ak^2, \quad (4)$$

where a and b are parameters to be estimated from the data fitting procedures.

Equations (1), (3), and (4) constitute the revised KF-ECC model together.

3.2. The Revised KF Model Based on Historical Deviation.

Since the rail transit passenger flow fluctuates dramatically and the magnitude is rather large, the forecasting process of KF model using passenger volume as a state variable directly is not very stable. Further analyses of passenger flows show that the deviation between real-time volume and the corresponding historical data is fairly smooth [37]. Therefore, the above-mentioned deviation is introduced into the KF model as the revised state variable to improve the accuracy and stability of the prediction. The revised KF-HD model is formulated as follows:

$$\mathbf{Q}(k) - \mathbf{H}(k) = [\mathbf{Q}(k-1) - \mathbf{H}(k-1)] + \mathbf{W}(k), \quad (5)$$

$$\mathbf{Q}^H(k) - \mathbf{H}(k) = [\mathbf{Q}(k) - \mathbf{H}(k)] + \mathbf{e}(k), \quad (6)$$

where $\mathbf{Q}^H(k)$ is the column vector form of historical passenger flow $Q^H(k)$ in the same time interval k and the same weekday during the last week. The most important issue is that $\mathbf{Q}^H(k)$ is different from $\mathbf{H}(k)$; that is, $\mathbf{Q}^H(k)$ is corresponding to the same weekday in the previous week, while $\mathbf{H}(k)$ is the average value of the historical data.

Equations (5) and (6) together constitute the revised KF-HD model, which is a basic KF formulation except for the state variable in a deviation form. Since $\mathbf{Q}^H(k)$ and $\mathbf{H}(k)$ are available from statistical data, one can get the real-time passenger flow $\mathbf{Q}(k)$ easily.

3.3. The Revised KF Model Based on Bayesian Combination and Nonparametric Regression.

Existing researches [31–34]

have proved the effectiveness of Bayesian combined approach in traffic flow forecasting. It is a weighted average method in fact, as shown below:

$$\mathbf{Q}(k) = \sum_{i \in I} \omega_i(k) \times \mathbf{Q}_i(k), \quad I = \{\text{KF}, \text{NR}\}, \quad (7)$$

where KF is the result from the KF model, NR is the result from the NR model, and ω_i is the weight of the KF or the NR model.

As stated before, the NR model is fairly applicable to uncertain and dynamic transportation systems, and many literatures have demonstrated its accuracy. Therefore, we introduce the NR method into the Bayesian combined model to further improve the prediction effects. Here, the K -nearest neighbor nonparametric regression (KNNNR) method is employed.

From (7), we can find out that, in the Bayesian combination framework, KF model or NR model may be strengthened or weakened by adjusting the weight ω_i . If we set ω_{KF} to zero, the KF model will be neglected from the combination. The same result will be derived for the NR model if we set ω_{NR} to zero. Actually, both weights will be adjusted dynamically according to the forecasting errors of two single models. The detailed adjustment mechanism will be illustrated in Section 4.

We further take the NR prediction as the control variable and introduce it into the KF model. Meanwhile, we combine the NR result in interval k with the KF result in interval $k-1$ through Bayesian combination method and integrate them into the state transition equation of the KF model. The revised formulation is shown below:

$$\mathbf{Q}(k) = \omega_{\text{KF}}(k) \cdot \mathbf{Q}_{\text{KF}}(k-1) + \omega_{\text{NR}}(k) \cdot \mathbf{Q}_{\text{NR}}(k) + \mathbf{W}(k), \quad (8)$$

where $\mathbf{Q}_{\text{KF}}(k)$ and $\mathbf{Q}_{\text{NR}}(k)$ are the column vector forms of $Q_{\text{KF}}(k)$ and $Q_{\text{NR}}(k)$, respectively, and other symbols are the same as before. The item $\omega_{\text{NR}}(k) \cdot \mathbf{Q}_{\text{NR}}(k)$ is the control variable of the state transition equation; that is, it reflects the contributions of NR model to the final prediction results.

Equations (8) and (2) constitute the revised KF-BCNR together. The main purpose of this revised KF model is to introduce more historical information and accurate results into the forecast process and to improve the accuracy and stability of the prediction.

Based on the adjusted algorithm of Bayesian weights and the results of the NR model, we can finally obtain the forecasted passenger flows.

4. Algorithms

4.1. Nonparametric Regression Algorithm. The NR algorithm mainly consists of five steps: the preparation of historical data, the generation of sample database, the definition of state vector, the searching of K -nearest neighbors, and the prediction function. The general algorithm flow is shown in Figure 1.

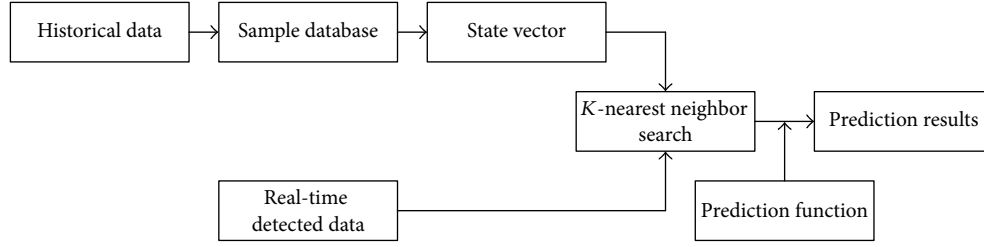


FIGURE 1: General flow of the NR algorithm.

Detailed algorithm is described as follows.

Step 1 (preparation of historical data). All historical detected data are prepared for the NR algorithm in this paper.

Step 2 (generation of the sample database). The prepared historical data are summarized into the sample database, which keeps updating with the forecast process and integrates both real-time data and historical data. The quality of the sample database greatly influences the performance of the NR model.

Step 3 (definition of state vector). Rail transit passenger flows are different from link traffic volumes; that is, there are no upstream or downstream links. However, when forecasting the station, some other stations near it will influence the arrival and distribution characteristics of its passenger flow. Therefore, we introduce the correlation analysis between target station and other stations. The number of correlative stations is determined by the correlation coefficient ρ_{AB} . Meanwhile, the state vector should include the passenger volumes of previous l intervals of the target station, where l is determined by the autocorrelation coefficient ρ_l with rank l .

Using $\{V_1^A, \dots, V_n^A\}$ to denote the time-series of passenger volumes during consequent n intervals of station A and $\{V_1^B, \dots, V_n^B\}$ to indicate the time-series of passenger volumes during consequent n intervals of station B, the correlation coefficient between stations A and B is formulated as

$$\rho_{AB} = \frac{\sum_{k=1}^n (V_k^A - \bar{V}^A)(V_k^B - \bar{V}^B)}{\sqrt{\sum_{k=1}^n (V_k^A - \bar{V}^A)^2 \sum_{k=1}^n (V_k^B - \bar{V}^B)^2}}, \quad (9)$$

where \bar{V}^A is the average of time-series $\{V_1^A, \dots, V_n^A\}$ and \bar{V}^B is the average of time-series $\{V_1^B, \dots, V_n^B\}$.

For the autocorrelation coefficient, we decompose the time-series of passenger volumes of the target station, $\{V_1, \dots, V_n\}$, into some subsequences with $n-l$ elements, that is, $\{V_1, \dots, V_{l+1}\}, \{V_2, \dots, V_{l+2}\} \dots \{V_{n-l}, \dots, V_n\}$, and then the autocorrelation coefficient is formulated as

$$\rho_l = \frac{\sum_{k=1}^{n-l} (V_k - \bar{V}_k)(V_{k+l} - \bar{V}_{k+l})}{\sqrt{\sum_{k=1}^{n-l} (V_k - \bar{V}_k)^2 \sum_{k=1}^{n-l} (V_{k+l} - \bar{V}_{k+l})^2}}. \quad (10)$$

Here, \bar{V}_k means the average of time-series of $\{V_k, \dots, V_{k+l}\}$.

Step 4 (searching of K -nearest neighbor). K -nearest neighbor search is to choose K -nearest data similar to current state vector and to predict the result of the next time interval based on the selected neighbors.

Euclidean distance is employed as the index to determine the K -nearest neighbor; that is,

$$d = \sqrt{\sum_{i=1}^I [V_i(k) - V_i^H(k)]^2 + \sum_{j=0}^l [V(k-j) - V^H(k-j)]^2}, \quad (11)$$

where I is the set of other stations correlated to the target station; $V_i(k)$ is the passenger volume of station i during interval k ; $V_i^H(k)$ is the historical data corresponding to $V_i(k)$; $V(k-j)$ is the passenger flow of the target station during interval $k-j$; $V^H(k-j)$ is the historical data corresponding to $V(k-j)$; d is the Euclidean distance.

Step 5 (prediction function). The prediction function is presented as in the following equation:

$$V(k+1) = \sum_{i=1}^K \frac{1/d_i}{d} V_i(k), \quad (12)$$

where K is the number of the most similar data serials, that is, the K -nearest neighbors; $d = \sum_{i=1}^K (1/d_i)$.

Using the above five steps, we can implement the NR algorithm and obtain the prediction results from the NR model. The above algorithm is coded using M language of the MATLAB platform.

4.2. The Sequential Kalman Filtering Algorithm. For the purpose of accuracy and efficiency, a sequential KF algorithm is employed to solve three revised KF models, which is illustrated in detail in our previous work [38]. This algorithm is also coded through the M language of the MATLAB software.

4.3. Bayesian Combination Algorithm. The key issue of Bayesian combination is weight of each submodel, which is decided logically according to the error comparisons of two single forecast methods.

Based on the historical prediction results and corresponding historical detection data, we can obtain the forecast

errors of the KF and the NR models, respectively. Here, the mean absolute percentage error (MAPE) is employed to denote forecast errors, as below:

$$\text{MAPE} = \frac{\sum_{k=1}^n (|\bar{Q}(k) - Q(k)| / Q(k))}{n} \times 100\%, \quad (13)$$

where $\bar{Q}(k)$ is the forecasted passenger flow during interval k , $Q(k)$ is the corresponding actual value, and n is the total number of time intervals.

Furthermore, we denote the historical MAPE of KF and NR models by EH^{KF} and EH^{NR} , respectively. The prior probabilities of choosing KF and NR models are then presented as

$$\Pr(H^{\text{KF}}) = \begin{cases} 1 - EH^{\text{KF}}, & (EH^{\text{KF}} < 1) \\ 0, & (EH^{\text{KF}} \geq 1), \end{cases} \quad (14)$$

$$\Pr(H^{\text{NR}}) = \begin{cases} 1 - EH^{\text{NR}}, & (EH^{\text{NR}} < 1) \\ 0, & (EH^{\text{NR}} \geq 1), \end{cases}$$

where $\Pr(\cdot)$ denotes a choice probability function; $\Pr(H^{\text{KF}})$ is the prior probability of choosing the KF model; $\Pr(H^{\text{NR}})$ is the prior probability of choosing the NR model. These two prior probabilities reflect the influences of historical forecasting errors.

To further incorporate the influences of current forecasting errors, we denote the current MAPE of the KF and the NR models by E^{KF} and E^{NR} , respectively. One must know that the current MAPEs are obtained based on the previous five time intervals; that is, they keep updating with the prediction process:

$$\Pr(F | H^{\text{KF}}) = \begin{cases} 1 - E^{\text{KF}}, & (E^{\text{KF}} < 1) \\ 0, & (E^{\text{KF}} \geq 1), \end{cases} \quad (15)$$

$$\Pr(F | H^{\text{NR}}) = \begin{cases} 1 - E^{\text{NR}}, & (E^{\text{NR}} < 1) \\ 0, & (E^{\text{NR}} \geq 1), \end{cases}$$

where $\Pr(F | H^{\text{KF}})$ and $\Pr(F | H^{\text{NR}})$ are the probabilities generating forecast F using the KF and the NR models, respectively.

Then, the posterior probabilities [33, 34] are formulated as

$$\Pr(H^{\text{KF}} | F) = \frac{\Pr(F | H^{\text{KF}}) \Pr(H^{\text{KF}})}{\Pr(F)},$$

$$\Pr(H^{\text{NR}} | F) = \frac{\Pr(F | H^{\text{NR}}) \Pr(H^{\text{NR}})}{\Pr(F)}, \quad (16)$$

$$\Pr(F) = \Pr(F | H^{\text{KF}}) \Pr(H^{\text{KF}}) + \Pr(F | H^{\text{NR}}) \Pr(H^{\text{NR}}),$$

where $\Pr(H^{\text{KF}} | F)$ and $\Pr(H^{\text{NR}} | F)$ are posterior probabilities of the KF and the NR models, respectively.

Based on (16), we finally obtain the weights of the KF and the NR models, as below:

$$\omega_{\text{KF}} = \frac{\Pr(F | H^{\text{KF}}) \Pr(H^{\text{KF}})}{\Pr(F | H^{\text{KF}}) \Pr(H^{\text{KF}}) + \Pr(F | H^{\text{NR}}) \Pr(H^{\text{NR}})}, \quad (17)$$

$$\omega_{\text{NR}} = \frac{\Pr(F | H^{\text{NR}}) \Pr(H^{\text{NR}})}{\Pr(F | H^{\text{KF}}) \Pr(H^{\text{KF}}) + \Pr(F | H^{\text{NR}}) \Pr(H^{\text{NR}})}.$$

Equations (7), (8), and (17) are integrated collectively as the revised KF-BCNR model.

5. Case Study

We collected the bus Smart Card Data (SCD) of line 13 of Beijing in the whole month of November 2013 and extracted the passenger volumes of 15 stations in every minute from such SCD information for a case study. According to the unified numbering rules of Beijing rail transit system, these 15 stations are named 21, 23, 25, 27, 29, 33, 35, 37, 39, 41, 43, 45, 47, 49, and 51, respectively. The operation period of line 13 is from 4:55 a.m. to 23:50 p.m. For application purpose, original data were aggregated to five minutes. Therefore, we totally have 228 time intervals. Passenger flows of station number 25 on November 28 (Thursday) were taken as the prediction target.

Using the above data, we implemented the KF model, the NR model, and the three proposed revised KF models and derived the prediction results of all five models, respectively.

5.1. Analyses of the NR Model. The state vectors are decided based on the correlation coefficient ρ_{AB} and the autocorrelation coefficient ρ_l , which are from time-series of passenger volumes of the target station and nearby stations, as shown in (9) and (10). Results show that the correlation coefficients between target station 25 and stations 21, 23, 27, and 49 all exceed 0.9; however, station 49 is excluded due to the relatively long distance from the target station. Therefore, the passenger flows of stations 21, 23, and 27 are taken as components of the state vector. Meanwhile, comparisons of the autocorrelation coefficients of the target station show that ρ_l is the biggest (0.86) when l equals 2.

The K -nearest neighbors are further determined by several forecasting experiments. Besides MAPE, three other evaluation indices are also employed to analyze the prediction errors, as below:

(1) MPE (mean percentage error):

$$\text{MPE} = \frac{\sum_{k=1}^n ((\bar{Q}(k) - Q(k)) / Q(k))}{n} \times 100\%. \quad (18)$$

(2) RMSE (root mean square error):

$$\text{RMSE} = \sqrt{\frac{\sum_{k=1}^n [\bar{Q}(k) - Q(k)]^2}{n}}. \quad (19)$$

(3) NRMS (normalized root mean square error):

$$\text{NRMS} = \frac{\sqrt{n \sum_{k=1}^n [\bar{Q}(k) - Q(k)]^2}}{\sum_{k=1}^n Q(k)} \times 100\%. \quad (20)$$

Other symbols in (18) to (20) are the same as above.

The error statistics of MAPE, MPE, RMSE, and NRMS in case of different K are summarized in Table 1.

From Table 1, one can find out that the general performance is the best while K equals 2. Therefore, K is determined as 2 in the K -nearest neighbor nonparametric regression model.

5.2. Prediction Results of the Three Revised KF Models. All information needed in the three revised KF models is extracted from the database. As stated before, the error correction coefficient λ in the revised KF-ECC model is determined by historical data fitting procedures:

$$\lambda = 0.010742k - 0.000045k^2. \quad (21)$$

Obviously, it is a quadratic parabola formulation.

In the revised KF-BCNR model, the historical data is necessary for the Bayesian weights. Here, information of November 21, the same Thursday during the previous week, is employed to get those weights.

Prediction results of the KF, NR, revised KF-ECC, revised KF-HD, and revised KF-BCNR models during the whole day are all reported in Table 2.

From Table 2, one can find out that all three revised KF models yield better results than the traditional KF model. In detail, introduction of the error correction coefficient makes the KF-ECC model outperform the original KF model. Employment of the Historical Deviation as state variable further improves the forecast accuracy of the KF-HD model. Integration of Bayesian combination and NR method yields the best performance for the KF-BCNR model. Meanwhile, the NR model is also rather accurate; however, its efficiency is not very satisfying for on-line applications.

To compare the performances of traditional models and three revised models during different periods, the evaluation indices during morning peak hours (7:00–9:00), nonpeak hours (11:00–13:00), evening peak hours (17:00–19:00), and the whole day (4:55–23:55) are further extracted and summarized in Table 3.

Graphical illustrations of these prediction results and errors during different periods are further described in Figures 2–9.

A further comparison of the prediction errors among all five models is illustrated in Figure 10. Here, the MAPE is employed to denote the forecasting error.

From the above predictions, one can find out the following results:

- (1) All the three revised KF models are fairly accurate for short-term rail transit passenger flows prediction. The revised KF-ECC model gets better results than the traditional KF model, due to the introduction

TABLE 1: Prediction error statistics of NR model.

K	MAPE	MPE	RMSE	NRMS
1	18.7%	8.2%	16.3	17.2%
2	19.7%	−3.4%	10.2	12.0%
3	21.4%	−8.6%	14.3	16.5%
4	24.0%	10.4%	21.5	25.2%
5	26.9%	−14.9%	27.4	29.1%

TABLE 2: Prediction error statistics of five models.

Model	MAPE	MPE	RMSE	NRMS
KF	38.8%	−33.5%	52.0	60.2%
NR	19.7%	−3.4%	10.2	12.0%
KF-ECC	27.8%	−14.9%	33.0	38.2%
KF-HD	20.5%	7.5%	11.5	13.3%
KF-BCNR	18.1%	4.1%	10.2	11.9%

TABLE 3: Prediction error statistics of five models during different periods.

Error indices	Models				
	KF	NR	KF-ECC	KF-HD	KF-BCNR
MAPE (%)					
Morning peak	35.5	10.4	12.9	10.2	8.2
Nonpeak	39.7	15.7	27.2	16.4	13.4
Evening peak	36.9	3.5	23.3	6.0	4.9
Whole day	38.8	19.7	27.8	20.5	18.1
MPE (%)					
Morning peak	−35.5	−3.3	−8.2	4.5	0.8
Nonpeak	−39.7	−9.3	−26.6	−0.8	−4.9
Evening peak	−36.9	−1.6	−23.3	2.6	0.6
Whole day	−33.5	−3.4	−14.9	7.5	4.1
RMSE					
Morning peak	33.4	11.3	13.6	9.9	7.6
Nonpeak	16.9	8.2	12.7	7.3	6.9
Evening peak	134.9	14.0	85.1	22.5	21.1
Whole day	52.0	10.2	33.0	11.5	10.2
NRMS (%)					
Morning peak	38.4	13.0	15.7	11.4	8.7
Nonpeak	44.4	21.5	33.3	19.1	18.1
Evening peak	38.8	4.0	24.5	6.5	6.1
Whole day	60.2	11.8	38.2	13.3	11.9

of the error correction coefficient. The revised KF-HD model further outperforms the KF-ECC model, because employing Historical Deviation as state variable improves its accuracy. Integrating Bayesian combination and the NR methods, the revised KF-BCNR model yields the best accuracy among all three revised KF models.

- (2) Concerning the capability of tracking the dynamic characteristics of real-time passenger flows, the three revised KF models also outperform the original KF method. Again, the revised KF-BCNR model

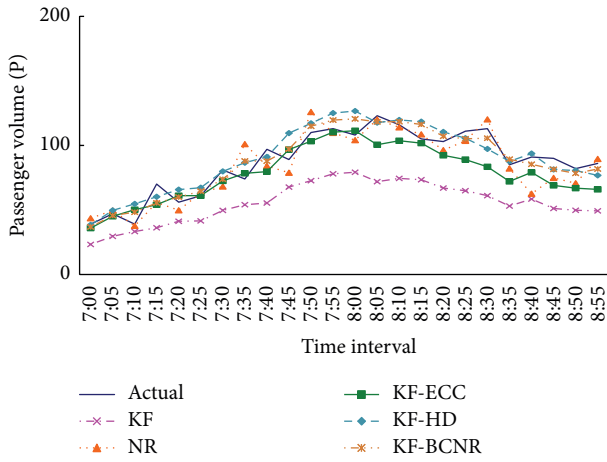


FIGURE 2: Prediction results in morning peak hours.

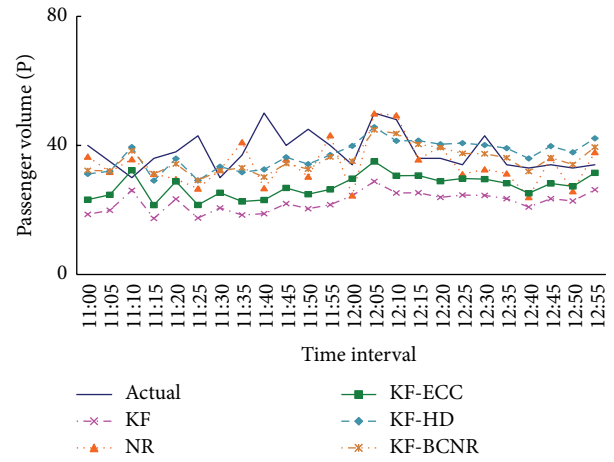


FIGURE 4: Prediction results in nonpeak hours.

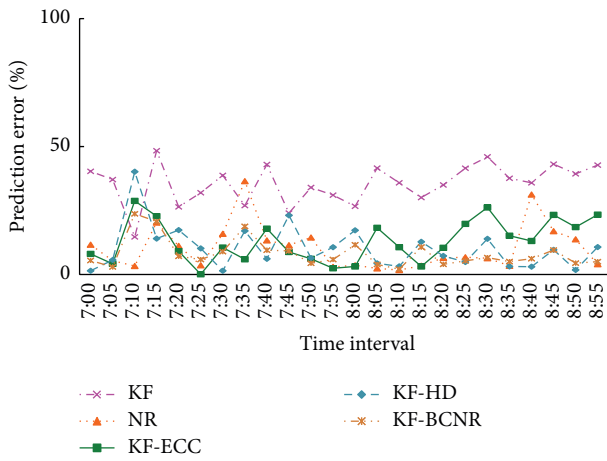


FIGURE 3: Prediction errors in morning peak hours.

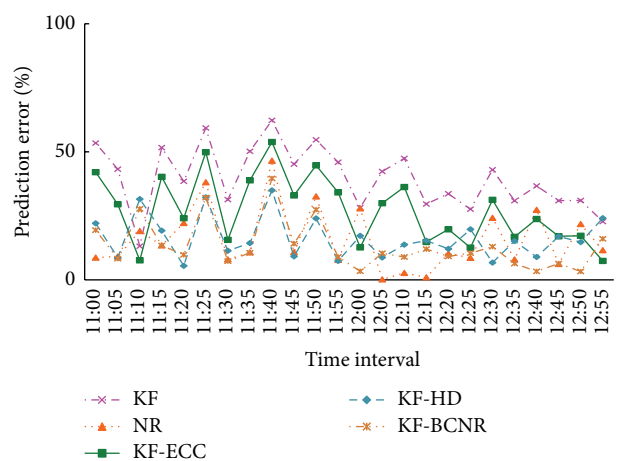


FIGURE 5: Prediction errors in nonpeak hours.

improved the stability significantly and yields the best result.

- (3) As a nonlinear regression method, the NR model gets much better results than the original KF model. It is even more accurate than the revised KF-ECC model in some cases. However, the revised KF-BCNR is still the most excellent model.
- (4) The comparisons among different periods show that the prediction performance during peak hours is much better than during nonpeak hours. The intrinsic reason is that the passenger volumes during peak hours are much bigger than those during nonpeak hours, and the fluctuations of passenger flows during peak hours are much weaker than those during nonpeak hours. Moreover, the much big magnitude of passenger volume during peak hours also reduces some error indices, for instance, MAPE, MPE, and NRMS, because of the sum of actual passenger flows in the denominator.
- (5) Prediction results during evening peak hours are the most accurate in all cases, with the MAPE at just 4.9%

and the NRMS at just 6.1%. The direct reason is that the passenger volume during this period is the highest and the most stable among all the time intervals.

- (6) Evaluation indices for the whole day are not very satisfying, because the passenger volumes during early morning and evening are very low and unstable, which can be seen from Figure 8. The very big errors corresponding to these time intervals in Figure 9 also indicate this phenomenon. These specific passenger flows greatly influence the prediction process and cause the increases of corresponding error indices.

Generally, all the three revised KF models are rather accurate and stable for on-line applications, especially during the very important peak hours.

6. Conclusions

This paper addresses three revised Kalman filtering models regarding short-term rail transit passenger flow prediction: the revised KF-ECC model, the revised KF-HD model, and

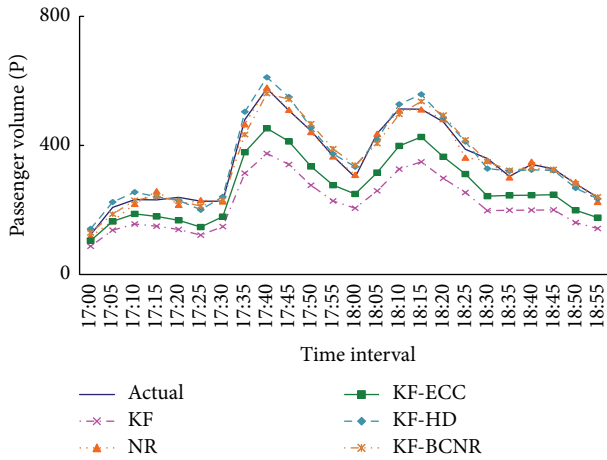


FIGURE 6: Prediction results in evening peak hours.

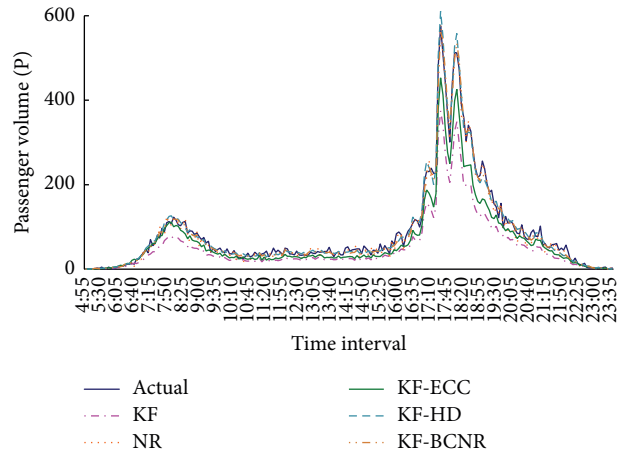


FIGURE 8: Prediction results in the whole day.

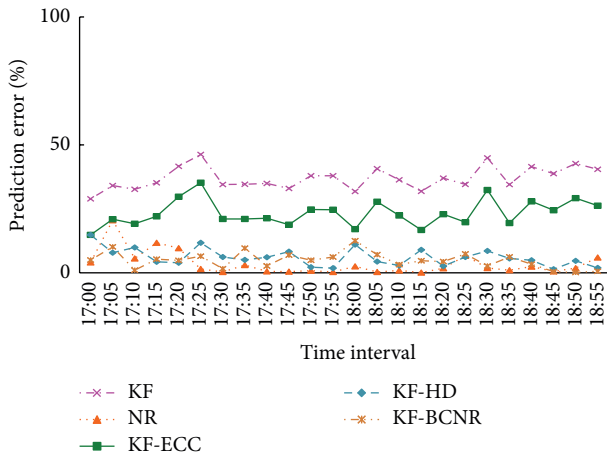


FIGURE 7: Prediction errors in evening peak hours.

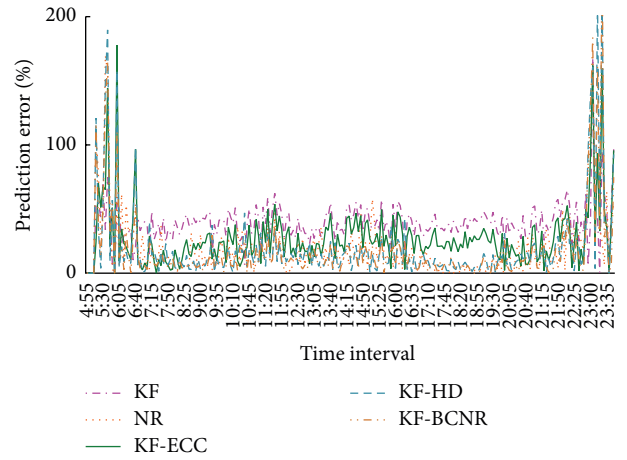


FIGURE 9: Prediction errors in the whole day.

the revised KF-BCNR model. We first present a revised KF-ECC model by introducing the historical prediction error into the measurement equation through an error correction coefficient. Since the original state variable fluctuates dramatically, we further employ the deviation between real-time passenger volume and corresponding historical data as a new state variable and derive a revised KF-HD model. For more accurate prediction, we integrate both the Bayesian combination technique and the nonparametric regression method into the traditional KF model and formulate a revised KF-BCNR model. The bus Smart Card Data of line 13 of Beijing during one-month period are collected for case study. The reported prediction results based on the practical data indicate that all three revised models are much more accurate and stable than traditional methods. Moreover, the revised KF-HD model outperforms the KF-ECC method, and the revised KF-BCNR model yields the best performance. Further comparisons among different periods show that predictions during peak hours are much more accurate than those during nonpeak hours, and forecast results during evening peak hours are the most excellent ones. Since peak hours are more important for rail transit operation and

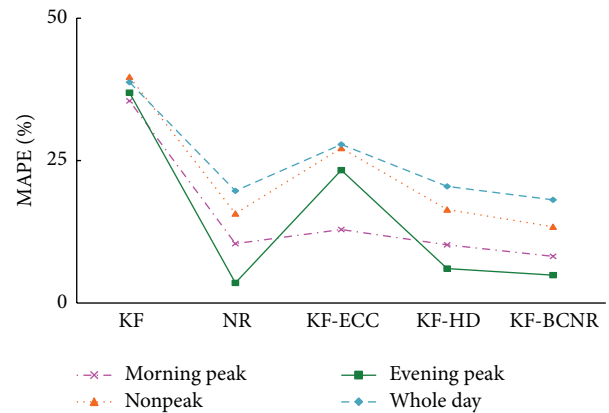


FIGURE 10: Comparison of prediction errors for different models and periods.

management, all three revised KF models proposed in this paper are accurate and stable enough for on-line applications.

Future potential research directions mainly consist of the following aspects. The first is to transform the three revised KF models to a short-term traffic flows forecast and to testify their applicability. The second is to further revise

the models and algorithms for applications in the whole rail transit system or large-scale road networks. The third is to explore the inherent interrelations among dynamic passenger volume, real-time urban travel demand, and rail network structure and to propose more logical prediction models based on dynamic travel demand analysis.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

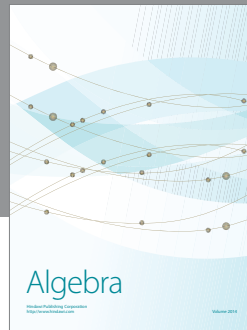
Acknowledgments

This research is supported by the National Natural Science Foundation of China Project (51578040, 51208024), Beijing Nova Programme (Z151100000315050), Beijing Natural Science Foundation Project (8162013), and the Importation and Development of High-Caliber Talents Project of Beijing Municipal Institutions (CIT&TCD201404071).

References

- [1] Y. Sun, G. Zhang, and H. Yin, "Passenger flow prediction of subway transfer stations based on nonparametric regression model," *Discrete Dynamics in Nature and Society*, vol. 2014, Article ID 397154, 8 pages, 2014.
- [2] Y. J. Stephanedes, P. G. Michalopoulos, and R. A. Plum, "Improved estimation of traffic flow for real-time control," *Transportation Research Record*, no. 795, pp. 28–39, 1981.
- [3] M. S. Ahmed and A. R. Cook, "Analysis of freeway traffic time-series data by using Box-Jenkins techniques," *Transportation Research Record*, no. 722, pp. 1–9, 1979.
- [4] B. M. Williams, P. K. Durvasula, and D. E. Brown, "Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models," *Transportation Research Record*, no. 1644, pp. 132–141, 1998.
- [5] B. L. Smith and M. J. Demetsky, "Short-term traffic flow prediction: neural network approach," *Transportation Research Record*, no. 1453, pp. 98–104, 1994.
- [6] L. Florio and L. Mussone, "Neural-network models for classification and forecasting of freeway traffic flow stability," *Control Engineering Practice*, vol. 4, no. 2, pp. 153–164, 1996.
- [7] H. J. Zhang, S. G. Ritchie, and Z.-P. Lo, "Macroscopic modeling of freeway traffic using an artificial neural network," *Transportation Research Record*, no. 1588, pp. 110–119, 1997.
- [8] M. S. Dougherty, H. R. Kirby, and R. D. Boyle, "The use of neural networks to recognise and predict traffic congestion," *Traffic Engineering & Control*, vol. 34, no. 6, pp. 311–314, 1993.
- [9] D. Park and L. R. Rilett, "Forecasting multiple-period freeway link travel times using modular neural networks," *Transportation Research Record*, no. 1617, pp. 163–170, 1998.
- [10] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach," *Transportation Research Part C*, vol. 13, no. 3, pp. 211–234, 2005.
- [11] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transportation Research Part B*, vol. 18, no. 1, pp. 1–11, 1984.
- [12] F. W. Cathey and D. J. Dailey, "A prescription for transit arrival/departure prediction using automatic vehicle location data," *Transportation Research Part C: Emerging Technologies*, vol. 11, no. 3–4, pp. 241–264, 2003.
- [13] S. Shekhar and B. M. Williams, "Adaptive seasonal time series models for forecasting short-term traffic flow," *Transportation Research Record*, no. 2024, pp. 116–125, 2007.
- [14] S. Yakowitz, "Nearest-neighbour methods for time series analysis," *Journal of Time Series Analysis*, vol. 8, no. 2, pp. 235–247, 1987.
- [15] M. Karlsson and S. Yakowitz, "Rainfall-runoff forecasting methods, old and new," *Stochastic Hydrology and Hydraulics*, vol. 1, no. 4, pp. 303–318, 1987.
- [16] G. A. Davis and N. L. Nihan, "Nonparametric regression and short-term freeway traffic forecasting," *Journal of Transportation Engineering*, vol. 117, no. 2, pp. 178–188, 1991.
- [17] B. L. Smith and M. J. Demetsky, "Traffic flow forecasting: comparison of modeling approaches," *Journal of Transportation Engineering*, vol. 123, no. 4, pp. 261–266, 1997.
- [18] R. K. Oswald, W. T. Scherer, and B. L. Smith, "Traffic flow forecasting using approximate nearest neighbor nonparametric regression," Research Report Uvacts-15-13-7, Center for transportation studies at the University of Virginia, Charlottesville, VA, USA, 2001.
- [19] B. L. Smith, B. M. Williams, and R. Keith Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 10, no. 4, pp. 303–321, 2002.
- [20] Y. Qi and B. L. Smith, "Identifying nearest neighbors in a large-scale incident data archive," *Transportation Research Record*, no. 1879, pp. 89–98, 2004.
- [21] M. D. Kindzerske and D. Ni, "Composite nearest neighbor nonparametric regression to improve traffic prediction," *Transportation Research Record*, no. 1993, pp. 30–35, 2007.
- [22] K. Huang, S. Chen, and Z. Zhou, "Research on a nonlinear chaotic prediction model for urban traffic flow," *Journal of Southeast University*, vol. 19, no. 4, pp. 410–414, 2003.
- [23] J. Lu and Z. Wang, "Prediction of network traffic flow based on chaos characteristics," *Journal of Nanjing University of Aeronautics and Astronautics*, vol. 38, no. 2, pp. 217–221, 2006.
- [24] Q. Meng and Y. Peng, "A new local linear prediction model for chaotic time series," *Physics Letters, Section A: General, Atomic and Solid State Physics*, vol. 370, no. 5–6, pp. 465–470, 2007.
- [25] J.-N. Xue and Z.-K. Shi, "Short-time traffic flow prediction using chaos time series theory," *Journal of Transportation Systems Engineering and Information Technology*, vol. 8, no. 5, pp. 68–72, 2008.
- [26] M.-B. Pang and X.-P. Zhao, "Traffic flow prediction of chaos time series by using subtractive clustering for fuzzy neural network modeling," in *Proceedings of the 2nd International Symposium on Intelligent Information Technology Application (IITA '08)*, pp. 23–27, Shanghai, China, December 2008.
- [27] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Statistics for Engineering and Information Science, Springer, New York, NY, USA, 2nd edition, 2000.
- [28] J. Ren, X. Ou, Y. Zhang, and D. Hu, "Research on network-level traffic pattern recognition," in *Proceedings of the IEEE 5th International Conference on Intelligent Transportation Systems*, pp. 500–504, Singapore, 2002.
- [29] 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.

- [30] J. Wang, X. Chen, and S. Guo, "Bus travel time prediction model with ν -support vector regression," in *Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems (ITSC '09)*, pp. 1–6, St. Louis, Mo, USA, October 2009.
- [31] W. Zheng, D.-H. Lee, and Q. Shi, "Short-term freeway traffic flow prediction: Bayesian combined neural network approach," *Journal of Transportation Engineering*, vol. 132, no. 2, pp. 114–121, 2006.
- [32] S. Dong, R. Li, L. G. Sun, T. H. Chang, and H. Lu, "Short-term traffic forecast system of Beijing," *Transportation Research Record*, no. 2193, pp. 116–123, 2010.
- [33] P. Jiao, T. Sun, and L. Du, "A bayesian combined model for time-dependent turning movement proportions estimation at intersections," *Mathematical Problems in Engineering*, vol. 2014, Article ID 607195, 8 pages, 2014.
- [34] P. Jiao, M. Liu, J. Guo, and T. Sun, "Bi-bayesian combined model for two-step prediction of dynamic turning movement proportions at intersections," *Advances in Mechanical Engineering*, vol. 2014, Article ID 439031, 9 pages, 2014.
- [35] Z. Yang, Q. Bing, C. Lin, N. Yang, and D. Mei, "Research on short-term traffic flow prediction method based on similarity search of time series," *Mathematical Problems in Engineering*, vol. 2014, Article ID 184632, 8 pages, 2014.
- [36] Z. Yang, D. Mei, Q. Yang, H. Zhou, and X. Li, "Traffic flow prediction model for large-scale road network based on cloud computing," *Mathematical Problems in Engineering*, vol. 2014, Article ID 926251, 8 pages, 2014.
- [37] K. Ashok and M. E. Ben-Akiva, "Alternative approaches for real-time estimation and prediction of time-dependent origin-destination flows," *Transportation Science*, vol. 34, no. 1, pp. 21–36, 2000.
- [38] P. Jiao and T. Sun, "Multiobjective traffic signal control model for intersection based on dynamic turning movements estimation," *Mathematical Problems in Engineering*, vol. 2014, Article ID 608194, 8 pages, 2014.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

