

XIAOBO ZHU, Ph.D. candidate^{1, 2}
E-mail: 1986zhuxb@tongji.edu.cn

JIANHUA GUO, Ph.D.²
(Corresponding Author)
E-mail: jg2nh@yahoo.com

WEI HUANG, Ph.D.²
E-mail: hhhwei@126.com

FENGQUAN YU, M.Sc.²
E-mail: 1003002915@qq.com

BYUNGKYU BRIAN PARK, Ph.D.³
E-mail: bpark@virginia.edu

¹ The Key Laboratory of Road and Traffic Engineering
(Ministry of Education), Tongji University
4800 Cao'an road, Jiading District,
Shanghai, China 201804

² Intelligent Transportation System Research Center,
Southeast University,
Si Pai Lou #2, Nanjing, P.R. China 210096

³ Department of Civil and Environmental Engineering,
University of Virginia
351 McCormick Rd, Charlottesville, VA 22904, USA

Traffic in the Cities
Preliminary Communication
Submitted: 2 Jan. 2017
Accepted: 23 Nov. 2017

REAL-TIME SHORT-TERM FORECASTING METHOD OF REMAINING PARKING SPACE IN URBAN PARKING GUIDANCE SYSTEMS

ABSTRACT

Short-term forecasting of the remaining parking space is important for urban parking guidance systems (PGS). The previous methods like polynomial equations and neural network methods are difficult to be applied in practice because of low accuracy or lengthy initial training time which is unfavourable if real-time training is carried out on adapting to changing traffic conditions. To forecast the remaining parking space in real-time with higher accuracy and improve the performances of PGS, this study develops an online forecasting model based on a time series method. By analysing the characteristics of data collected in Nanjing, China, an autoregressive integrated moving average (ARIMA) model has been established and a real-time forecasting procedure developed. The performance of this proposed model has been further analysed and compared with the performances of a neural network method and the Markov chain method. The results indicate that the mean error of the proposed model is about 2 vehicles per 15 minutes, which can meet the requirements for general PGS. Furthermore, this method outperforms the neural network model and the Markov chain method both in individual and collective error analysis. In summary, the proposed online forecasting method appears to be promising for forecasting the remaining parking space in supporting the PGS.

KEY WORDS

parking guidance system; remaining parking space; time series method; short-term forecasting method; neural network;

1. INTRODUCTION

Parking guidance systems (PGS) can provide drivers with parking information and routing guidance to public parking lots in a multiple-stage system and hence improve the efficiency of regional parking facilities and relieve the urban parking problems [1]. The en-trip parking lot selection decision made by the drivers depends on many factors, such as the distance from the parking lot to their destinations, parking cost, spare parking space, etc. Due to limited parking space, the parking problem is becoming serious in metropolitan areas, especially in the central areas of big cities. If there are not enough facilities to meet the parking demand, the drivers should have to pay more attention to the information on whether there are spare parking spaces near their destinations. Indeed, parking space information is the basis of PGS serving drivers [2]. Only when there are spare spaces in parking lots, the drivers will choose these places to park and pay attention to further services, such as route guidance, parking space reservation, and other related services. Therefore, processing and dissemination of the remaining parking space information is vital for the PGS.

In the planning and design periods of a PGS, the stratifications of the parking guidance areas and parking information are necessary because of the large scope of the guidance region [3]. When drivers receive high-level information, there would be a long distance

from the parking lots to their destinations and the remaining parking space may vary during the course of the driver approaching the parking lot which they have selected. Therefore, if only real-time information about the remaining parking space is published to drivers, it may happen that when they arrive at the chosen parking lot, the actual condition is different from the information they had received previously, especially in the central area or at peak times. To avoid this phenomenon and provide accurate parking information to the drivers, the remaining parking space should be forecasted into short-term future and disseminated to the drivers. The length of the short-term period depends on the real-time operation of the parking facilities and the time the drivers need to travel from the place where they receive the parking space information to the chosen parking lot. Furthermore, improving the accuracy and reliability of PGS is beneficial to control and manage the urban parking traffic in macroscopic aspect and increase the efficiency of urban traffic operations [4, 5]. Therefore, it is necessary to develop an effective method for forecasting the remaining parking space accurately to make the guidance information accurate and reliable, and such ability of predicting the remaining parking space is a precondition for the PGS operating effectively.

Some previous studies tried to establish effective forecasting methods for the remaining parking space forecasting based on experiences or modelling the relationship between the influencing factors and the number of the remaining parking spaces. Jia [6] presented two methods to forecast the available parking space shown in variable message signs of PGS, i.e., the simple empirical method and the random variable model, and the random variable model was found to provide better results. These models established the relationships between the parking space and the influencing factors. David and Keller [7] presented a method to forecast on-street remaining parking space. Based on the sociodemographic and parking-specific data, this method took into account the current and future events which influence the general demand for the parking space. Since there are no automatic detectors in on-street parking space, this method is based on historical data collected by periodic manual counts with low frequency. Therefore, if the parking space changes frequently, the accuracy and validity of this model would be affected. Hu [8] proposed the empirical-based models to forecast effective parking space of a parking facility. However, since many kinds of factors may affect the remaining parking spaces, including internal factors like types of car parks, parking cost and position of car parks, and uncertain random factors such as weather and crashes, and for different parking facilities, the main factors may be very different, and the abovementioned methods cannot involve all the potential factors. Furthermore, these models

need to be adjusted continuously according to the actual traffic and parking conditions, which might consume long updating time on data processing. In this case, when drivers arrive at the chosen parking lot, the actual number of the remaining parking spaces may be different from the information they had received, which will reduce the reliability of PGS.

For many years, the time series models, which are established based solely on the regularity of the data series and make forecasts into the future, have been applied to traffic prediction problems [9-11]. Especially in short-term prediction of traffic flow, the Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) model is applied under many conditions and some valuable results have been achieved [12-14]. Guo and Deng [15] established an ARIMA model for short-term traffic flow forecasting, and results indicated that the model is practical and feasible. Kumar and Vanajakshi [16] developed a seasonal ARIMA model to forecast the traffic flow with limited data. Lin [17] combined wavelet analysis and ARIMA model to reduce the forecasting error and improve the feasibility of ARIMA model. These methods can avoid the difficulty of incorporating many influencing factors, and the only information needed for the modelling process would be the data series itself, which will simplify and hence improve the reliability of the short-term forecasting approached built thereupon [18]. Therefore, some researchers developed time series methods to forecast the remaining parking spaces [19-21]. However, few studies applied the ARIMA model for the prediction of the remaining parking spaces. Yang, Liu and Wang [19] applied the neural network model into the prediction of available parking spaces. They found that although large sample sizes could reduce the prediction error, a lot of the calculation time was costly and the real-time requirement could not be satisfied. Further studies were conducted by SUN and Cui [22], CHEN et al. [23] to improve the neural network models. However, the problems about calculation time and real-time requirement were not completely solved. Ji, Wang and Deng [20] proposed short-term forecasting method for the remaining parking spaces using wavelet analysis and Markov models based on the characteristics of the occupied parking spaces. Ji, Tang and Guo [21] further combined the largest Lyapunov exponents and wavelet neural network into both one-step and multi-step forecasting. This new method was more accurate than the traditional wavelet neural network models. However, the real-time forecasting process was not studied. In addition, multi-step forecasting accumulates the prediction errors and hence reduces the prediction accuracy.

These methods for the prediction of the remaining parking spaces mentioned above are difficult to be widely applied in practice because of low accuracy or lengthy initial training time which is unfavourable if

real-time training is carried out on adapting to changing traffic conditions. Therefore, targeting the need for developing a real-time remaining parking space forecasting approach and recognizing the periodicity and similarity observed in the real world remaining parking space data, the Box-Jenkins time series method derived from economics and statistics is introduced into the parking guidance systems in this study so as to explore the regularity existing in the data series for conducting short-term forecasting.

The main objective of this study is to develop a method to forecast the remaining parking spaces in real-time more accurately and hence improve the performances of PGS. To establish the forecasting model, the real-world data of Central Emporium Mall Parking Lot in Nanjing, China have been collected. The contributions of this study include three aspects: (1) three patterns of changing characteristics of real-world remaining parking space data are analysed for supporting the development of a time series model, including monthly pattern, weekly pattern and daily pattern; (2) based on the analysis, an Autoregressive Integrated Moving Average (ARIMA) model is established and an online forecasting approach based on the model is developed to forecast in real-time the number of remaining parking spaces; (3) the accuracy and reliability of this online forecasting approach is verified by performance analysis and comparison with two typical forecasting methods used in the previous studies.

2. PROPOSED REAL-TIME FORECASTING METHOD

Based on the specification of a time series model and the discussion on the time series modelling procedure, a real-time short-term remaining parking space forecasting method is proposed in this section.

2.1 Time series model specification

Time series is a series of data that observed values of a statistics are arranged according to time, showing the pattern of this statistics changing with time. Time series models can be applied to describe the characteristics and the tendency of the historically observed data series and make forecast into the future [18].

The influencing factors are not considered in time series forecasting models. Instead, the relationship between the historical and the future data is the only consideration. An important assumption of time series models is that the trend of the data series would last into the future. Therefore, time series method has great potential in short-term forecasting through exploring the trending structure. In PGS, the number of the remaining parking spaces is a series of data arranged according to time which can be forecasted by the time series method and hence the information on

the remaining parking spaces can be published to drivers in real-time during their trips. Conventionally, many time series models have been applied in the economics field. For example, Yule proposed the Autoregressive (AR) model to forecast the change of market and Box-Jenkins method was developed further with strong theoretical basis and a structured modelling process. The Box-Jenkins approach is still the main method of time series analysis that is applied in more and more research fields [24].

There are four common time series models in the family of the Box-Jenkins time series models [25, 26]: Autoregressive (AR) model, Moving Average (MA) model, Autoregressive Moving Average (ARMA) model, and Autoregressive Integrated Moving Average (ARIMA) model. The first three models are all applied in modelling and forecasting the stationary time series. In real world, many data series are unstable or non-stationary with unfixed mean values and variances, which cannot be modelled by those three models. However, non-stationary time series can be transformed to stationary series by applying the differencing operator appropriately, resulting in the ARIMA model. Therefore, it can be seen that the first three models are simply special cases of the ARIMA model.

Given time series X_t , the mathematical expression of ARIMA (p, d, q) model is Equation 1.

$$(1 - B)^d X_t = \mu + \frac{\theta(B)}{\Phi(B)} a_t \quad (1)$$

where p is the order of the autoregressive part; q is the order of the moving-average process; d is the order of the differencing; t is the indexed time; μ is the mean term; a_t is a random error; B is the backshift operator which is calculated as $BX_t = X_{t-1}$; $\theta(B)$ is the moving-average operator which is $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$; $\Phi(B)$ is the autoregressive operator which is $\Phi(B) = 1 - \Phi_1 B - \dots - \Phi_p B^p$.

In fact, the ARIMA models assume linear relationship between the data series, and in this paper, this linear structure assumption based forecasting approach will be compared with those based on non-linear assumptions, i.e. the neural network model.

2.2 Time series modelling procedure

In linear time series models, AR, MA, and ARMA models can be considered as special cases of the ARIMA model. So their modelling processes are similar. In this section, the modelling process of the ARIMA forecast model is introduced, including the model identification stage, model estimation and diagnostic checking stage, and forecasting stage [25].

Model identification stage

The main purpose of this stage is to test the stationarity of time series and identify candidate ARIMA models. In this stage, the autocorrelation functions

and descriptive statistics of time series are calculated first. The stationarity is tested based on the results in reference [27]. If the time series is non-stationary, the differencing will be conducted. Testing the stationarity and differencing are repeated until differenced series reaches stationarity. Then the order of the autoregressive part and the moving-average process are identified by observing the characteristics of the partial autocorrelation function and the inverse autocorrelation function [28]. According to these identification rules, one or more candidate ARIMA models can be identified for further investigation.

Estimation and diagnostic checking stage

The parameters of suggested candidate models will be estimated in this stage and diagnostic checking will be conducted to determine which model is the best fit to the series in question [29, 30]. The following activities will be done in this stage: (1) estimating the parameters by maximum likelihood estimation or least squares estimation; (2) significance tests to drop insignificant parameters; (3) comparing the adequacy of model to determine the optimal model by the conventional information criteria: Akaike Information Criterion (AIC) [31] or Bayesian Information Criterion (BIC) [32]; (4) white noise test for the residuals to show whether the autocorrelation structure has been removed from the original series. In this stage, the most parsimonious model is preferable and a final ARIMA model will be selected for the succeeding forecasting step.

Forecasting stage

In this stage, the ARIMA model established in the first two stages is applied to forecast the future values of the time series. In doing so, the forecasting equation would be constructed using the final identified parsimonious model and the estimated parameters, then, the past values can be supplied into the forecasting equation so that forecasts could be computed for the future time steps.

However, it is necessary to point out that usually one step ahead forecasting is conducted in this step, and considering the need for forecasting the

remaining parking space continuously over time, it is important for the short-term forecasting to be conducted in real-time to reflect the changing parking information. In this way, a real-time forecasting scheme should be designed to operate the time series model in an online fashion as described below.

2.3 Real-time forecasting framework development

In real world, the information on the remaining parking spaces changes constantly and hence forecasts generated using a fixed model may be inaccurate. Therefore, the time series used to do the forecast should be updated in real-time and a better forecasting method is to forecast the data at the next time interval and update the time series model continuously. Bearing this in mind, and through wrapping up the aforementioned Box-Jenkins time series modelling procedure, a real-time forecasting process is developed and shown in *Figure 1*, with four steps described as below: *Step 1*: determine the data series at time t ; *Step 2*: establish the forecasting ARIMA model by identification, estimation and diagnostic checking; *Step 3*: forecast the number of remaining parking spaces for time $t+1$ and publish the information to the users; *Step 4*: update the data series at time $t+1$ and return to Step 2.

By collecting real-time remaining parking space data and updating the data series continuously, the parameters of the forecasting ARIMA model can be estimated and amended in real time. Therefore, the number of the remaining parking spaces can be forecasted in real time. This process can be implemented using SAS software. Considering the established efficiency of SAS-based ARIMA modelling procedures, this process requires little time in updating real-time data and forecasting, i.e., less than one minute in the empirical case of this paper. Therefore, the proposed time series model-based online forecasting method can forecast the number of remaining parking spaces in real time, which is important for improving the reliability of the PGS.

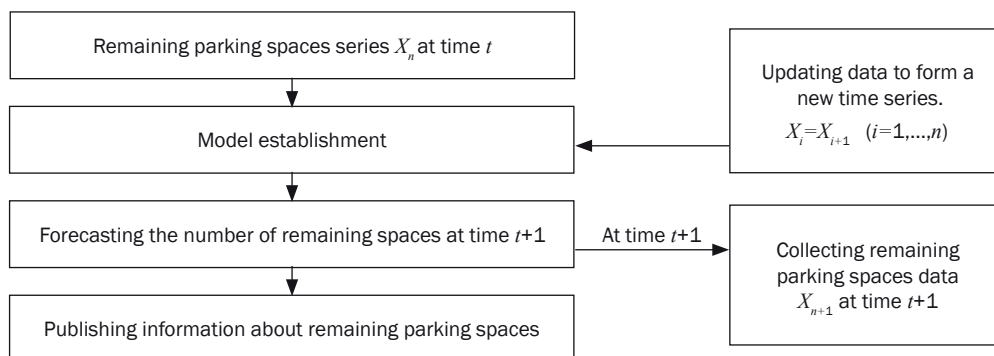


Figure 1 – Real-time forecasting procedure

This real-time forecasting procedure is a one-step forecasting procedure. By adjusting the time interval of data series, the length of prediction interval can be changed depending on the real-time operation of parking facilities and the time the drivers need to travel from the places where they have received parking space information to the chosen parking lots.

3. EMPIRICAL STUDY

In this section, the real-world parking data from the Central Emporium Mall Parking Garage in Nanjing are used to develop and validate the online remaining parking space forecasting model. The characteristics of the remaining parking spaces are analysed for developing the time series model, and the prediction accuracy of the proposed online forecasting model is also verified and compared with the one of the neural network method.

3.1 Data

The Central Emporium Mall is located in the main business centre of Nanjing, China. The parking demand of this area is high, such that parking supply is often below the demand. It is common in the peak period that a large number of vehicles waits for the remaining parking spaces of the garages in this area. Since the Central Emporium Mall is one of the main businesses in this area, it is significant to relieve the parking problem by improving the service efficiency of its garage.

The Central Emporium Mall garage has a total of 180 parking spaces. An advanced electronic parking management system is used to manage the parking lots. The vehicle access control, guidance in parking lots and payment are managed by the system. The arriving and departing data of every vehicle can be collected from the system. In this study, one-month data were collected for October 2010, and the data were aggregated at a 15-minute interval. The parking data collected from the electronic parking management system were the arriving and departing times of vehicles. By pre-processing the data, the number of the remaining parking spaces in every fifteen minutes was obtained.

3.2 Characteristics of remaining parking spaces

Three patterns of changing characteristics of the remaining parking spaces have been investigated, including monthly pattern, weekly pattern and daily pattern. The monthly pattern shows notable periodicity as in *Figure 2a*. The periodic curve has a similar trend as sine curves. In addition, the difference between the changing trends in workdays and weekends are remarkable. Therefore, it can be concluded that the

cycle of the remaining parking spaces appears to be weekly. The weekly pattern of the remaining parking spaces is shown in *Figure 2b* from 11 to 17 October 2010. It shows that the parking spaces are in short supply every afternoon, especially during weekends. *Figure 3* shows the daily patterns in a typical weekday and a typical weekend, respectively. For the weekend, due to high demand on parking, the number of the remaining parking spaces decreases to zero quickly after the mall opens, and the garage remains nearly full when the mall is open. The remaining parking spaces are also in shortage on weekdays while it shows slightly increased vacancies than on the weekends in the garage.

From the remaining parking space series analysis above, the changing characteristics of the workdays and weekends are notably different and the trend of the remaining parking spaces is cyclic in every week. Therefore, to establish the regularity existing in the data series and to conduct accurate short-term forecasting, the sample data used to establish the time series model should include more than one week's data.

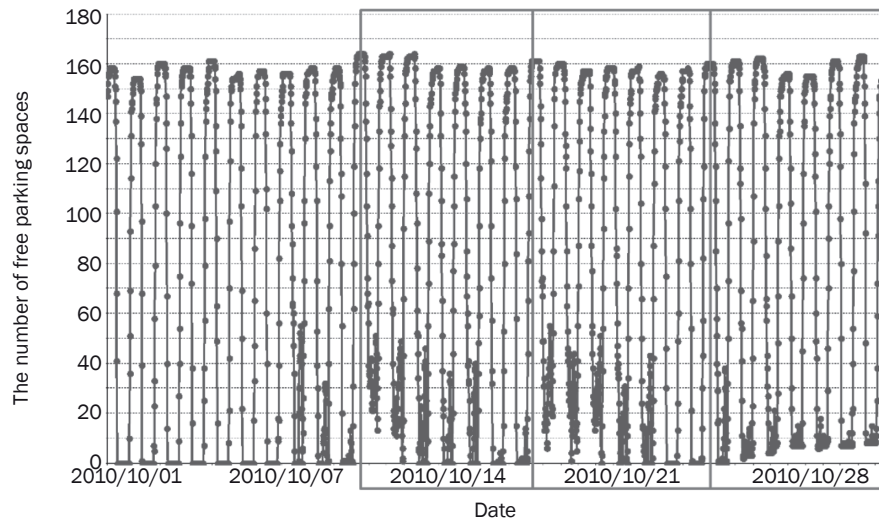
In this study, the remaining parking space data series from 1 to 30 October 2010 are used to establish and calibrate the forecasting model, while the data on 31 October 2010 are used to analyse the performance of the model.

3.3 Time series model identification

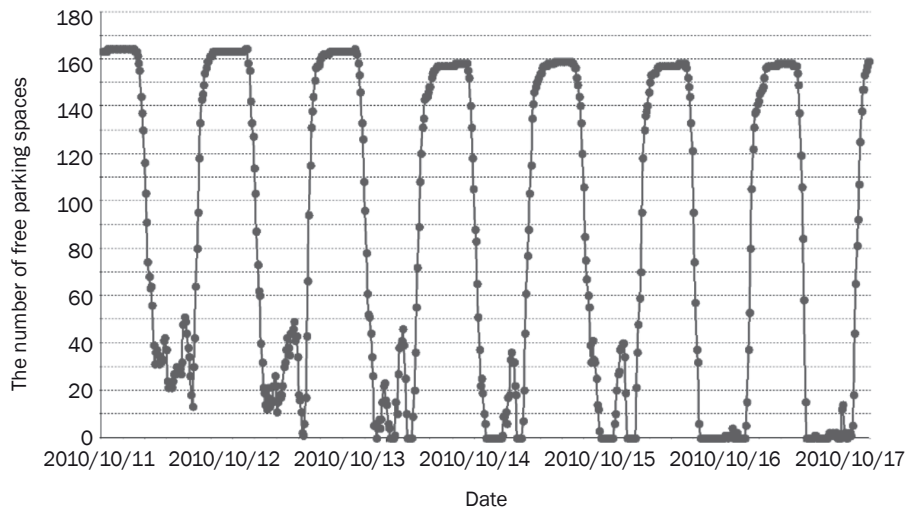
PROC ARIMA of SAS software identifies time series models by following the three stages of the Box-Jenkins method: model identification, parameter estimation and forecasting. The stationarity of the original remaining parking space data is analysed using autocorrelation function in *Table 1*. To show the autocorrelation coefficients clearly, the latest 25 values are listed. The

Table 1 – Autocorrelation of original remaining parking space data

Lag	Correlation	Lag	Correlation
0	1.000	13	0.531
1	0.993	14	0.485
2	0.977	15	0.439
3	0.953	16	0.393
4	0.921	17	0.347
5	0.886	18	0.300
6	0.845	19	0.253
7	0.802	20	0.205
8	0.758	21	0.157
9	0.713	22	0.109
10	0.668	23	0.061
11	0.622	24	0.013
12	0.575	25	-0.036



a) Monthly pattern



b) Weekly pattern

Figure 2 – Characteristics of the remaining parking spaces

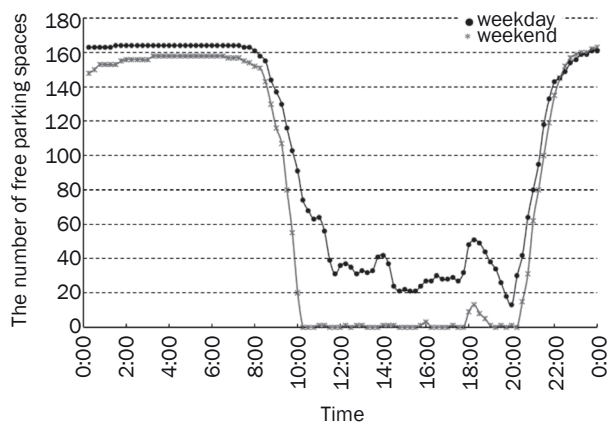


Figure 3 – Daily characteristics of the remaining parking spaces

Lag column indicates the time difference used in computing the correlation, e.g., the correlation for lag 0 are those for the series itself, and the value of 0.993 in the

correlation column for Lag means that the correlation of the remaining parking spaces between adjacent time lags is 0.993. On observing Table 1, the correlations of the remaining parking spaces are significant and decrease slowly over time. Therefore, it can be concluded that the original remaining parking space data series has significant autocorrelation structure, indicating the necessity of modelling this structure to conduct short-term forecasting accordingly. Moreover, the slowing decaying autocorrelation function might also indicate a certain level of non-stationarity in the original remaining parking space series.

White noise test of the original series is illustrated in Table 2. The results show that p -value for the test is less than 0.0001 and the autocorrelations of the series are high. So the white noise hypothesis is rejected strongly, which is consistent with the previous findings of significant autocorrelation structure existing in the remaining parking space series.

Table 2 – Autocorrelation check for white noise

To lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	9999.99	6	< 0.0001	0.993	0.977	0.953	0.922	0.886	0.845
12	9999.99	12	< 0.0001	0.802	0.758	0.713	0.668	0.622	0.576
18	9999.99	18	< 0.0001	0.531	0.485	0.439	0.393	0.347	0.300
24	9999.99	24	< 0.0001	0.253	0.205	0.157	0.109	0.061	0.013

Table 3 – Autocorrelation of difference series

Lag	Correlation	Lag	Correlation
0	1.000	6	0.215
1	0.768	7	0.125
2	0.669	8	0.073
3	0.540	9	0.040
4	0.417	10	0.015
5	0.316		

Since original remaining parking space series is non-stationary, differencing is conducted and a differenced series is obtained. Autocorrelation function of differenced series shown in Table 3 indicates that the correlations decrease rapidly to near zero. Therefore, the differencing operator is necessary for processing the series.

To identify appropriate ARMA models to the differenced series, the partial and inverse autocorrelation functions are needed, and the values of p and q can be determined by the characteristics of the partial and inverse autocorrelations. If the partial or inverse autocorrelation values are consistently close to zero when lag is larger than a certain threshold number, p or q will be assigned as this threshold value. In this case, the values of p and q can be both selected to be 2, as shown in Table 4.

Table 4 – Partial autocorrelation and inverse autocorrelation of difference series

Partial autocorrelation		Inverse autocorrelation	
Lag	Correlation	Lag	Correlation
1	0.768	1	-0.319
2	0.194	2	-0.183
3	-0.060	3	-0.014
4	-0.081	4	0.038
5	-0.030	5	-0.038

However, the autocorrelation functions computed from these data only approximate the true functions. So the selected p and q might not exactly match the “true” model. Therefore, in this case, the range of p and q are both designated from 1 to 5, and the final model will be determined based on the information criterion in the next stage.

3.4 Time series model parameter estimation and diagnostic checking

In this stage, the orders of the ARIMA models will be selected and the parameters of the ARMA models are estimated. For each pair of p and q , the estimation and diagnostic checking are conducted with the information criteria (AIC and BIC) shown in Table 5.

Table 5 – Goodness-of-Fit for each TIME SERIES model

p	q	AIC	BIC	p	q	AIC	BIC
1	1	17,347.75	17,365.64	3	4	17,291.69	17,339.41
1	2	17,306.17	17,330.03	3	5	17,287.47	17,341.16
1	3	17,304.84	17,334.67	4	1	17,293.54	17,329.33
1	4	17,304.14	17,339.93	4	2	-*	-*
1	5	17,287.86	17,329.62	4	3	17,234.49	17,282.21
2	1	17,322.13	17,345.99	4	4	17,288.73	17,342.41
2	2	17,292.35	17,322.18	4	5	17,287.89	17,347.54
2	3	17,231.72	17,267.51	5	1	17,295.54	17,337.29
2	4	17,292.36	17,334.12	5	2	17,233.56	17,281.29
2	5	17,285.80	17,333.52	5	3	-*	-*
3	1	17,295.77	17,325.60	5	4	17,286.94	17,346.59
3	2	17,293.05	17,328.85	5	5	17,288.87	17,354.49
3	3	17,296.11	17,337.86				

* Unavailable data. Because of the problem of convergence, parameters cannot be estimated.

On observing *Table 5*, when $p=2$ and $q=3$, the values of AIC and BIC are the smallest. So ARMA (2,3) can be selected as the optimal model for the differenced series. Since the original series is differenced only once, ARIMA (2,1,3) is identified as the model fitting the remaining parking space series in this study, with the estimated model listed in *Equation 2*. Note that this model is an offline model in terms of ARIMA (2,1,3), serving as the initial forecasting equation in the proposed online forecasting method, and the model parameters will be updated accordingly with the incoming of future remaining parking space data.

$$(1 - B)X_t = -0.01421 + \frac{1 - 1.2273B + 0.34206B^2 - 0.11308B^3}{1 - 1.82419B + 0.83456B^2} a_t \quad (2)$$

3.5 Performance analysis and comparison

As noted, the remaining parking space data from 1 to 30 October 2010 have been used to establish the ARIMA model. The number of remaining parking spaces on 31 October 2010 is forecasted and compared with the collected data to analyse the accuracy of the proposed method. In addition, the results of a neural network forecasting method and a Markov chain forecasting method are compared with the proposed online model to verify the superiority of the proposed method. Note that the accuracy analysis includes two aspects, i.e., disaggregated error analysis at every time interval and collective error analysis of forecast results. The performance measures of the error analysis at every time interval are absolute error (AE) and absolute percentage error (APE), while the performance measures for the collective error analysis are mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) [10, 33].

The neural network forecasting method

Neural network is a theory which is inspired by the central nervous system of biology. A neural network is generally presented as a system of interconnected neurons which exchange information among each other. The connections have numeric weights that can be tuned based on past experiences [34, 35]. Back propagation neural network (BP neural network) is the most widely applied neural network. In this study, a BP neural network is used to forecast the number of the remaining parking spaces for comparing with the proposed method. For the BP network model, three layers of input layer, hidden layer, and output layer are included, and the gradient descent method was applied to calculate the minimum value of the error function. The weights of the connections are tuned according to the negative gradient direction of the error function until the error is lower than a predetermined threshold value of 0.001 or the training epochs reach a

predetermined value of 1,000,000 [36]. The training process of a BP neural network follows those presented in [36, 37]. In this paper, the training process is implemented in MATLAB neural network toolbox, and the number of neurons in the hidden layer is determined as three. The convergence curve of training is shown in *Figure 4*. When the error is less than 0.001, the training is stopped and a corresponding neural network model is determined. After training the BP model using the remaining parking space data from 1 to 30 October 2010 (i.e., the training dataset), the number of the remaining parking spaces on 31 October 2010 (i.e., the test dataset) can be forecasted. The forecasting process is also implemented in MATLAB neural network toolbox.

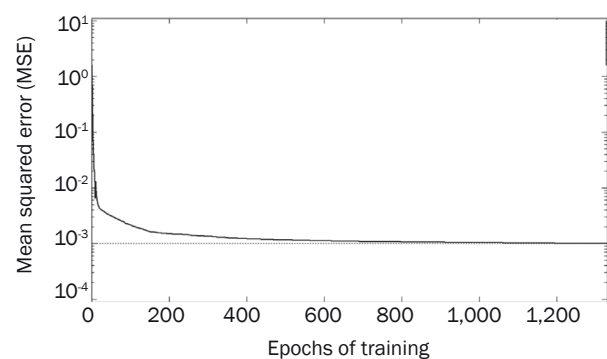


Figure 4 - The convergence curve of training neural network

Overfitting is a notable problem during the training process of the neural network method. The error on the training dataset is driven to a very small value, but when new data are presented to the network, the error is large [38]. Therefore, overfitting of the BP neural network would have negative effects on the performance of forecasting. In MATLAB neural network toolbox, the default method for addressing the overfitting problem is early stopping technique in which the dataset is divided into three subsets: the training set, the validation set and the test set. The error on the validation set is monitored during the training process. In the initial phase, the error normally decreases. If the validation error begins to increase for a specified number of iterations, the training will be stopped, and the weights and biases at the minimum of the validation error will be returned. The default data division is that 60% of the samples are assigned to the training set, 20% to the validation set, and 20% to the test set. Actually, if the number of parameters in the network is much smaller than the total number of points in the training dataset, there is little or even no chance of overfitting [39].

In this study, the default early stopping method is implemented to avoid overfitting. Overfitting does not appear in the network training and an optimal BP

neural network is returned when the error is lower than the predetermined threshold value (0.001), shown in Figure 4.

The Markov chain forecasting method

The Markov property which is named after the Russian mathematician Andrey Markov describes the transformation between different states of a system.

A Markov chain is a stochastic process with the Markov property which is named after the Russian mathematician Andrey Markov defining serial dependence only between adjacent periods. It describes the transformation between different states of a system and assumes that the future state of the system only depends on the current state [40]. According to the transition probability between different system states and the current state of the system, the future state can be forecasted, which is the theoretical basis of the Markov chain forecasting method.

The common forecasting process is described as below [40]:

Step 1: collected data series of the system;

Step 2: divide states of the system. State division is the foundation of Markov chain forecasting. The threshold values applied to divide the state depend on the sample size of the data set. When there are fewer data, the system state should be less. And with the increase in data amount the system state can be increased.

Step 3: calculate the transition probability between different system states.

Define $E=\{E_1, E_2, \dots, E_n\}$ as the set of system states, and f_{ij} as the frequency of transition from state E_i to state E_j , where $E_i, E_j \in E$. Matrix $F=\{f_{ij}\}$ is the transition frequency matrix of the system. The transition probability p_{ij} from state E_i to state E_j can be calculated as:

$$p_{ij} = \frac{f_{ij}}{\sum_{j=1}^n f_{ij}} \tag{3}$$

By calculating f_{ij} and p_{ij} , the transition probability matrix $P=\{p_{ij}\}$ can be obtained.

Step 4: forecast the future state of the system. Based on the transition probability matrix and current state of the system, future state can be forecasted.

In this study, data from 1 to 30 October 2010 are used for dividing states of the remaining parking spaces. According to the maximum and minimum number of the remaining parking spaces (165 and zero, respectively), the data series are divided into 33 states evenly. The multiples of 5 are the threshold values of these states.

In the forecasting step, the threshold value of future state is considered as the number of remaining parking spaces which is forecasted by the Markov chain forecasting method.

Disaggregated performance across time

Figure 5 shows the comparison between the collected remaining parking space data, together with the forecasted series using the proposed online forecasting method, the BP neural network method and the Markov chain method. On observing Figure 5, the proposed approach works desirably in that the forecasted series can track the observed series, especially when the remaining parking space changes smoothly. In contrast, for both the neural network method and the Markov method, the forecasting errors are notable even in low demand period, indicating that the proposed method outperforms the neural network approach and the Markov chain approach.

To show the detailed forecasting performances of the three approaches, AE and APE of these three methods are shown in Figures 6a and 6b, respectively.

On observing Figure 6a, most absolute errors by the proposed method are less than two vehicles per 15 minutes while the errors between 9:00 a.m. and 11:00 a.m. are larger than the errors in other periods. As shown in Figure 3b, the time around 9:00 a.m. is

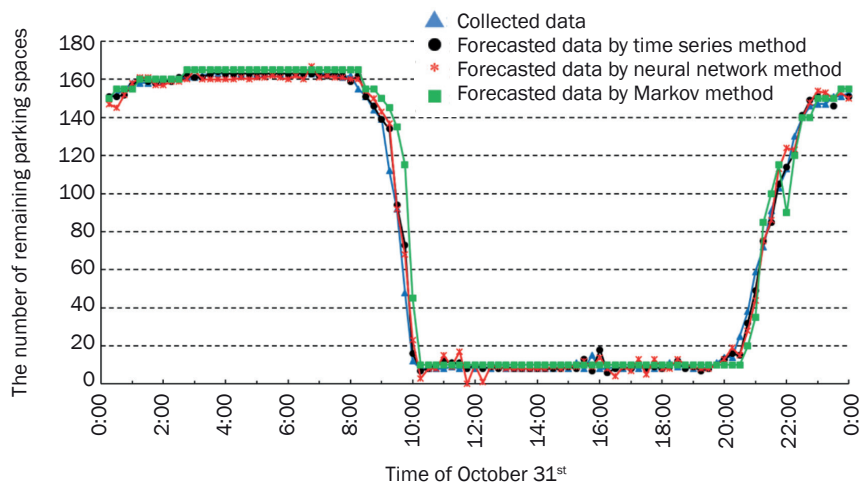
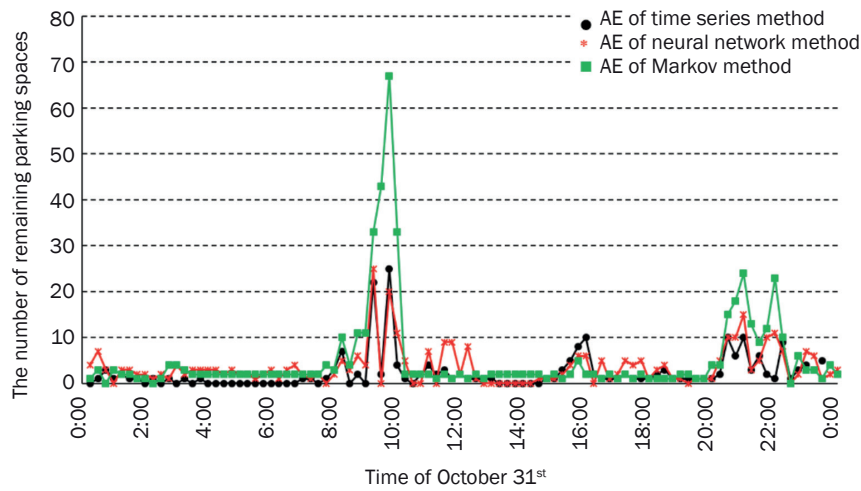
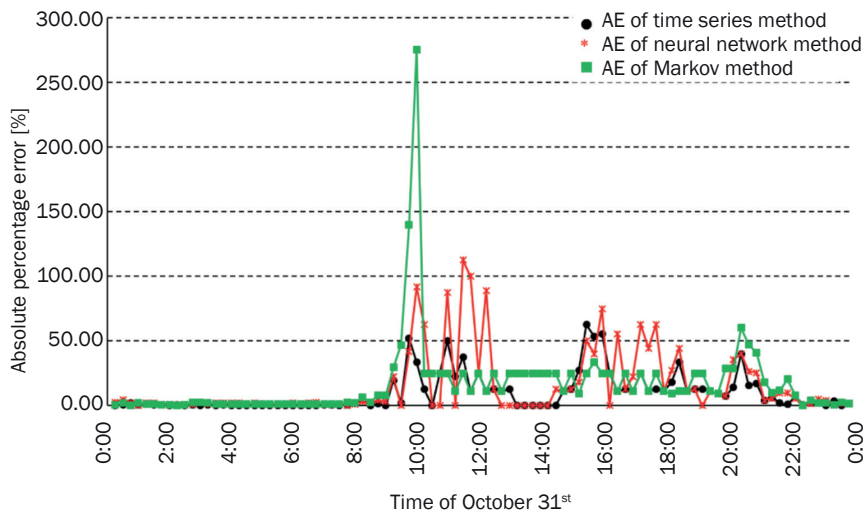


Figure 5 – Comparison between observed data and forecasted data



a) Absolute error



b) Absolute percentage error

Figure 6 – Error analysis at every time interval

the beginning of peak time and the remaining parking space reduced to the minimum level at about 11:00 a.m. Many vehicles park in the Central Emporium Mall during this short period, which makes the number of the remaining parking spaces difficult to be forecasted. Similar tendency appears between 8:00 p.m. and 10:00 p.m. in which the number of the remaining parking spaces also changes quickly since it is near the closing time of the mall. For the BP neural network method, many absolute errors are more than five vehicles per 15 minutes and fluctuate notably. For the Markov chain method, the absolute errors are even more than ten vehicles per 15 minutes when the number of the remaining parking spaces changes frequently, with the maximal absolute error being near 70 vehicles per 15 minutes.

On observing Figure 6b, the absolute percentage errors (APE) of the proposed method are smaller than with other two methods at most time intervals.

However, the APEs of all methods are large in the peak period. It is because the collected data in peak hours are small so that even a small difference may cause a large value of APE.

In summary, the proposed real-time forecasting method based on the time series model can be used to forecast the number of the remaining parking spaces, and the proposed method outperforms the conventional non-linear neural network model and the Markov chain forecasting model in terms of absolute error and absolute percentage error, indicating the validity of the proposed model.

Collective performances

The collective performances of the proposed method, the neural network method and the Markov chain method are investigated and compared in this section using performance measures of MAE, MAPE, and RMSE, calculated by Equations 4, 5 and 6.

$$MAE = \frac{1}{n} \sum_{t=1}^n |X_t - X'_t| \tag{4}$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - X'_t}{X_t} \right| \cdot 100\% \tag{5}$$

$$RMSE = \frac{1}{n} \sqrt{n \sum_{t=1}^n (X_t - X'_t)^2} \tag{6}$$

where X_t is the collected data of the remaining parking spaces at time t ; X'_t is the forecasted data at time t ; n is the sample size.

Based on the equations above, the values of MAE, MAPE, and RMSE for the three methods are shown in Table 6. All the collective errors of the proposed method are smaller than other two methods. Therefore, these results also show that the proposed method is better than the neural network method and the Markov chain method for forecasting the number of remaining parking spaces.

Table 6 – MAE, MAPE and RMSE for the two methods

Method	MAE	MAPE	RMSE
Proposed Method	2.22	9.12%	4.47
BP Neural Network	3.58	15.55%	5.45
Markov Method	5.16	16.78%	10.91

In summary, the results of both disaggregate and collective error analysis show that the proposed method has desirable performances on forecasting the number of the remaining parking spaces. In addition, through performance comparisons, the proposed method outperforms the conventional neural network method and the Markov chain method.

3.6 Applicability analysis of the proposed method

The general purpose of the Parking Guidance System is to guide drivers to a suitable parking lot in which there are vacant spaces. Some studies revealed that the most needed information to the drivers is the number of remaining parking spaces and direction to the facility [3, 19]. In some information dissemination strategies, if the number of the remaining parking spaces of a parking lot is lower than the threshold value, such as 5% of the total number of the parking spaces, “full” information will be published to drivers [24]. In Tapiola, Finland, this threshold value varies from 1% to 10%, depending on the distance between the variable message signs and the parking lot. The greater the distance, the larger the threshold value. In addition, to prevent rapid switch between “space” and “full”, larger threshold values are used when information moves back from “full” to “space” [2].

In practice, the Parking Guidance System generally provides parking information to drivers in multiple stages. Wide-area or regional parking information is published in high-level stages. In these stages, variable message signs are installed at the periphery of roads entering and leaving the whole guidance area or sub-areas, which are far away from the parking lots [1], and the threshold values are generally large due to long distances between message signs and the parking spot, accepting large forecasting errors in such cases. In this study, the mean absolute error of the proposed method is around two vehicles per 15 minutes, i.e., around 1.11% of the total number of parking spaces, which can be acceptable in most stages of PGS.

The real-time short-term forecasting method of the remaining parking spaces proposed in this study can conduct the forecasting process in less than one minute, has better performances than the conventional neural network method and the Markov chain method and can satisfy the demand of information publishing of PGS. Therefore, the proposed method in this study has great potential in meeting the demand of PGS and it is suitable to be applied in the real world to improve the accuracy and reliability of PGS.

4. CONCLUSIONS

The remaining parking space forecasting is one of the key technologies for improving the performances of parking guidance systems. In this study, based on analysing the characteristics of the remaining parking space data in the Central Emporium Mall garage in Nanjing, China, a real-time short-term forecasting method is proposed for forecasting the remaining parking space data.

On observing the monthly pattern, weekly pattern and daily pattern of the real-world remaining parking space series, the data have shown notable periodicity, and accordingly, the Box-Jenkins time series method is applied to model the number of the remaining parking spaces.

Following the Box-Jenkins method, an ARIMA (2,1,3) model is established with the results of diagnostic checking showing that this model is the optimal fitting to the remaining parking space data series. In addition, a real-time short-term forecasting framework is developed based on updating the ARIMA model over time. Although the forecasting procedure is a one-step procedure, the length of prediction interval can be changed by adjusting the time interval of data series to satisfy different forecasting requirements.

Using the real-time forecasting approach, the forecasting accuracy is analysed and compared with the neural network approach and the Markov chain method. In individual error analysis, most absolute errors (AE) of the proposed method are less than two

vehicles. In contrast, many AEs of the neural network approach are larger than five vehicles and fluctuate notably, while AEs of the Markov chain method are larger than ten vehicles during peak hours, which would reduce the reliability of PGS. In collective error analysis, the mean absolute error (MAE) and root mean square error (RMSE) of the proposed method are 2.22 and 4.47, respectively, while the MAEs of the neural network method and the Markov chain method are 3.58 and 5.16, respectively and the RMSEs of these two methods are 5.45 and 10.91, respectively. These results show that the proposed real-time forecasting method has better performances than both the neural network method and the Markov chain method in forecasting the number of the remaining parking spaces, which can improve the performances of PGS.

Future studies are needed to improve the forecasting accuracy under conditions when the number of the remaining parking spaces changes rapidly. Under these conditions, information about the remaining parking space is more important for drivers. In addition, for better accounting for the uncertainties associated with the forecasting, the second order analysis of the forecasting errors should be conducted in which prediction intervals can be constructed to reveal in detail the reliability of the forecast. This will further improve the reliability of the PGS system developed upon the short-term remaining parking space forecasting.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China under Grants No. 61573106 and No.51238008.

朱晓波^{1,2}

E-mail: 1986zhuxb@tongji.edu.cn

郭建华 (通讯作者)²

E-mail: jg2nh@yahoo.com

黄卫²

E-mail: hhhwei@126.com

于丰泉²

E-mail: 1003002915@qq.com

BYUNGKYU BRIAN PARK³

E-mail: bpark@virginia.edu

¹ 同济大学道路与交通工程教育部重点实验室 (教育部)
上海市嘉定区曹安公路4800号

² 东南大学智能运输系统研究中心
南京市四牌楼2号

³ Department of Civil and Environmental Engineering,
University of Virginia
351 McCormick Rd, Charlottesville, VA 22904, USA

停车诱导系统实时剩余泊位短时预测方法

摘要

停车场剩余泊位短时预测对停车诱导系统 (PGS) 应用效果具有重要意义。以往研究中基于多项式、神经网络

等模型的预测方法由于精度较低或需要较长的计算时间,难以应用于实时变化的现实交通环境。本文提出了一种基于时间序列模型的实时预测方法,以便准确地、实时地预测停车场剩余泊位数,从而改进停车诱导系统应用效果。通过分析南京市一处停车场剩余泊位变化特征,建立了ARIMA短时预测模型并提出了实时预测过程。本文进一步分析了模型的适用性和精度,并将其与以往研究中使用的神经网络预测模型和马尔科夫链预测模型进行对比。结果表明,以15分钟为预测步长的情况下,ARIMA模型平均预测误差为两个泊位,满足停车诱导系统应用的精度要求。同时,ARIMA模型在预测精度和计算时间方面均强于神经网络模型和马尔科夫链模型。因此,本文提出的剩余泊位实时预测方法可以有效地、准确地预测短时剩余泊位数,为停车诱导系统实际应用提供有力支撑。

关键词

停车诱导系统; 剩余泊位数; 时间序列分析方法; 短时预测方法; 神经网络

REFERENCES

- [1] Kurogo H, Takada K, Akiyama H. Concept of a Parking Guidance System and Its Effects. In: *The Shinjuku Area-configuration, Performance, and Future Improvement of System. Proceedings of the 6th International Vehicle Navigation and Information Systems Conference, 30 July - 2 Aug 1995, Washington, USA.*
- [2] Ristola T. Parking guidance system in Tapiola. *Proceedings of the 6th International Conference on Road Traffic Monitoring and Control, 1992, London, England.*
- [3] Spencer ME, West J. Parking Guidance System in San Jose, California. *Transportation Research Record.* 2004;1886: 34-39.
- [4] Yoo S, Chong PK, Kim T. PGS: Parking Guidance System Based on Wireless Sensor Network. *The 3rd International Symposium on Wireless Pervasive Computing, 7-9 May 2008, Santorini, Greece.* IEEE; 2008. p. 218-222.
- [5] Idris MYI, Tamil EM, Noor NM, Razak Z, Fong KW. Parking Guidance System Utilizing Wireless Sensor Network and Ultrasonic Sensor. *Information Technology Journal.* 2009;8(2): 138-146.
- [6] Jia JP. *Statistics.* Beijing: China Renmin University Press; 2000.
- [7] David A, Keller H. Event-driven Modeling of On-street Parking Probability. *Proceedings of the 8th World Congress on ITS, 2001, Sydney, Australia.*
- [8] Hu JJ. [Research on the Key Technologies of Parking Guidance System]. Dissertation, Jilin: Jilin University, 2006.
- [9] Caicedo F, Blazquez C, Miranda P. Prediction of Parking Space Availability in Real Time. *Expert Systems with Applications.* 2012;39(8): 7281-7290.
- [10] Guo F, Krishnan R, Polak J. A Computationally Efficient Two-stage Method for Short-term Traffic Prediction on Urban Roads. *Transportation Planning and Technology.* 2013;36(1): 62-75.
- [11] Cyprich O, Konecny V, Kilianova K. Short-term Passenger Demand Forecasting Using Univariate Time Series Theory. *Promet - Traffic & Transportation.* 2013;25(6): 533-541.
- [12] Zhang L, Li XY, Shi HJ. Study of Short-term Traffic Flow Forecasting Algorithms Based on ARIMA

- Model. *Journal of Zhengzhou University of Light Industry*. 2008;23(4): 89-92.
- [13] Lippi M, Bertini M, Frasconi P. Short-term Traffic Flow Forecasting: An Experimental Comparison of Time-series Analysis and Supervised Learning. *IEEE Transaction Intelligent Transportation System*. 2013;14(2): 871-882.
- [14] Brockwell P, Davis R. *Introduction to time series and forecasting*. Springer; 2011.
- [15] Guo X, Deng FQ. Short-term Prediction Of Intelligent Traffic Flow Based On BP Neural Network And ARIMA Model. *International Conference on E-Product E-Service and E-Entertainment*, 7 - 9 November, Henan, China. IEEE; 2010. p. 1-4.
- [16] Kumar SV, Vanajakshi L. Short-term Traffic Flow Prediction Using Seasonal ARIMA Model with Limited Input Data. *European Transport Research Review*. 2015;7(3): 21.
- [17] Lin J. Study on the Prediction of Urban Traffic Flow Based on ARIMA Model. *3rd International Conference on Engineering Technology and Application (ICETA 2016)*. 2016: 418-422.
- [18] Pandit SM and Wu SM. *Time Series and system analysis with application*. New York: John Wiley and Sons, Inc.; 1983.
- [19] Yang ZS, Liu HH, Wang XY. The Research on the Key Technologies for Improving Efficiency of Parking Guidance System. *Intelligent Transportation Systems*. 2003;2: 1177-1182.
- [20] Ji YJ, Wang W, Deng W. Available Parking Space Occupancy Change Characteristics and Short-term Forecasting Model. *Journal of Southeast University (English Edition)*. 2007;23(4): 604-608.
- [21] Ji YJ, Tang DN, Guo WH. Forecasting Available Parking Space with Largest Lyapunov Exponents Method. *Journal of Central South University*. 2014;21: 1624-1632.
- [22] Sun Y, Cui ZM. Research and Implementation of Forecasting Model for Parking Guidance System Based on Neural Network. *Computer Application and Software*. 2007;24(11): 124-126.
- [23] Chen Q, Yan KF, Wang RT, Mo YK. Parking Space Information Prediction Based on Phrase Construction and Elman Neural Network. *Journal of Tongji University (Natural Science)*. 2007;28 (4): 413-420.
- [24] Thompson RG, TakadK, Kobayakawa S. Optimization of Parking Guidance and Information Systems Display Configurations. *Transportation Research, Part C*. 2001;9(1): 79-85.
- [25] Box GEP, Jenkins GM, Reinsel GC. *Time series analysis: forecasting and control*. New York: Prentice Hall PTR Upper Saddle River; 1976.
- [26] Chen P. *Applied mathematical statistics*. Beijing: China Machine Press; 2008.
- [27] Dickey, DA, Fuller WA. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*. 1979;74(366): 427-431.
- [28] Ljung GM, Box GEP. On a Measure of Lack of Fit in Time Series Models. *Biometrika*. 1978;65(2): 297-303.
- [29] Ansley C, Newbold P. Finite Sample Properties of Estimators for Autoregressive Moving Average Models. *Journal of Econometrics*. 1980;13(2): 159-183.
- [30] Davidson J. Problems with the Estimation of Moving Average Models. *Journal of Econometrics*. 1981;16(3): 295-310.
- [31] Akaike H. A New Look at the Statistical Model Identification. *Transaction on Automatic Control*. 1974;19(6): 716-723.
- [32] Schwarz G. Estimating the Dimension of a Model. *Annals of Statistics*. 1978;6(2): 461-464.
- [33] Guo JH, Williams BM. Real-time Short Term Traffic Speed Level Forecasting and Uncertainty Quantification Using Layered Kalman Filters. *Transportation Research Record*. 2010;2175: 28-37.
- [34] Hagan MT, Demuth HB, Beale M. *Neural network design*. Boston: PWS Publishing Co.; 1996.
- [35] Smith BL, Demetsky MJ. Traffic Flow Forecasting: Comparison of Modeling Approaches. *Journal of Transportation Engineering*. 1997;123(4): 261-266.
- [36] Zheng WZ, Lee DH, Shi QX. Short-term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach. *Journal of Transportation Engineering*. 2006;132(2): 114-121.
- [37] Gao XP, Cong S. Comparative Study on Fast Learning Algorithms of BP Networks. *Control and Decision*. 2001;16(2): 167-171.
- [38] Minns AW, Hall MJ. Artificial Neural Networks as Rainfall-runoff Models. *Hydrological Sciences Journal*. 1996;41(3): 399-417.
- [39] Demuth HB, Beale M. *Neural network toolbox - for use with MATLAB*. Natick: The MathWorks, Inc.; 2004.
- [40] Ji YJ, Gao LP, Chen XS, Guo WH. Strategies for Multi-step-ahead Available Parking Spaces Forecasting Based on Wavelet Transform. *Journal of Central South University*. 2017;24(6): 1503-1512.