

Research Article

Space-Time Hybrid Model for Short-Time Travel Speed Prediction

Qi Fan ¹, Wei Wang ², Xiaojian Hu,² Xuedong Hua,² and Zhuyun Liu³

¹Jiangsu Key Laboratory of Urban ITS, Southeast University, Si Pai Lou #2, Nanjing 210096, China

²School of Transportation, Southeast University, Si Pai Lou #2, Nanjing 210096, China

³Zhuhai Institute of Urban Planning & Design, Mei Hua Dong Road #302, Zhuhai 519000, China

Correspondence should be addressed to Wei Wang; wangwei_transtar@163.com

Received 21 February 2017; Accepted 23 January 2018; Published 25 February 2018

Academic Editor: Gabriella Bretti

Copyright © 2018 Qi Fan 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-time traffic speed forecasting is a significant issue for developing Intelligent Transportation Systems applications, and accurate speed forecasting results are necessary inputs for Intelligent Traffic Security Information System (ITSIS) and advanced traffic management systems (ATMS). This paper presents a hybrid model for travel speed based on temporal and spatial characteristics analysis and data fusion. This proposed methodology predicts speed by dividing the data into three parts: a periodic trend estimated by Fourier series, a residual part modeled by the ARIMA model, and the possible events affected by upstream or downstream traffic conditions. The aim of this study is to improve the accuracy of the prediction by modeling time and space variation of speed, and the forecast results could simultaneously reflect the periodic variation of traffic speed and emergencies. This information could provide decision-makers with a basis for developing traffic management measures. To achieve the research objective, one year of speed data was collected in Twin Cities Metro, Minnesota. The experimental results demonstrate that the proposed method can be used to explore the periodic characteristics of speed data and show abilities in increasing the accuracy of travel speed prediction.

1. Introduction

Driven by the need to promote Intelligent Transportation Systems (ITS) and traffic safety management, short-time future travel speed predicting is a crucial issue that attracted a number of studies. Common forecasting methods mainly can be divided into five categories: methods based on traditional statistical theory, intelligent model methods based on knowledge discovery, methods based on nonlinear system theory, methods based on hybrid model, and other forecasting methods. Concrete methods include regression method, neural network methods, wavelet network models, time series model, support vector regression methods, Kalman filtering methods, exponential smoothing methods, gray system model, trend extrapolation method, and artificial intelligence.

Travel speed data often shows periodic trends, with one to two minimum value peaks in one weekday; however, such a trend may not be similar from Monday to Sunday because of the different characteristics of the day trip within a week.

Observing the cyclical changes in the regulation of the speed, one can grasp the daily variation characteristics, which could provide the planning and design indicators and basic services. Dendros [1] considered traffic as a combination of 17 periodic components and residual parts and as fitting the trend with Fourier series. Fei et al. [2] present a dynamic linear model to predict short-term travel time: travel time is the sum of the median of historical travel times, time-varying random variations in travel time, and evolution error model. Liu [3] presents that there is an intrinsic mode function in speed, and this part is predicted by ARIMA. Zhang et al. [4] present that periodic variation of traffic flow been analyzed by the statistical volatility model. It is important to note that periodic trend analysis cannot give the speed change in small scale, which means the prediction of small range float for speed becomes particularly important.

The residual part is the difference between the real-time series data and the periodic trend. With comparison and analysis of a great amount of data, the residual part shows an obvious fluctuation, but the volatility is stable within

a certain range. For such series, autoregressive integrated moving average (ARIMA) method has a unique advantage at the time of prediction. ARIMA model is one of the most widely used regression techniques. Its applications in freeway traffic forecasting can be traced back to 1979 [5]. When using the ARIMA model for traffic flow prediction [6], difference processing is performed on the data if it is a nonstationary sequence and has a certain growth or decline trend. If there has a heteroscedasticity in the data, the nonstationary sequence is required to be smoothed with the difference model until the autocorrelation function and the partial correlation function of the processed data are not significantly different from zero.

Lund [7] used R language to simulate the ARIMA process, with a great result. However, Karlaftis and Vlahogianni [8] point out that the ARIMA model lacks the ability to capture long memory properties and does not jointly treat the mean and variance.

Research on the speed prediction at the present stage focuses primarily on characteristics of time variation on a single point on the road, considers changing speed with time, lacks spatial analysis of the target position, and ignores the influence of predicted results from the travel speed change of adjacent links. However, links in the road network are not isolated; speed change of target position follows its own change rule in the dimension of time and is affected simultaneously by upstream and downstream traffic conditions. The speed of the upstream section can be transmitted to the downstream section through the road with similar distribution characteristics, and the traffic state of the downstream will also react to the upstream. Adding spatial correlation analysis can effectively detect the occurrence and influence of nearby plugging points and can reflect the impact of emergencies on speed, might warn of upstream and downstream traffic issues, and reduce the congestion and network congestion caused by emergencies in a timely manner through management means paralysis. At first, van Lint [9] established a state-space neural network (SSNN) model to predict travel times directly from the data of adjacent section. Pan et al. [10] developed a spatial and temporal dynamic model with stochastic cell transmission. Zou et al. [11] put forward a hybrid model combined with spatial analysis and several time analyses to compare and analyze the results.

Most of methods encounter the problem of lack of accuracy or reliability when used separately in travel speed prediction, and so we proposed a hybrid model to properly combine them. Using the real data to test and verify results of periodic trend and ARIMA, ARIMA method, spatial analysis regression model, and the hybrid model we proposed, the comparison indexes of predictive effects like RAME, MAE, and MAPE all indicate that our hybrid model can predict short-time travel speed accurately and reliably.

2. Methodology

The signal forecasting method for short-time travel speed has proved to be difficult in providing accurate information. We established a hybrid model combining the results of time prediction and spatial prediction so that the real

speed changes of freeways can be more accurately simulated.

2.1. Periodic Trend Analysis. Traffic speed usually changes daily. The cyclical change of travel speed is obvious, especially within a workday. Tang et al. [12] discussed the periodic characteristics of speed in detail. The periodic trend is a significant feature in dealing with speed data and should be considered first in analyzing speed change rules. Periodic analysis [7] is adopted to investigate the cyclicity in the daily traffic data and is also effective in analyzing travel speed rules.

From the perspective of regression analysis, change of travel speed appears a kind of trend of cyclical fluctuations; therefore, wave theory is feasible during the analysis. Fourier series express the periodic trend in traffic speed, which is driven by trigonometric functions. For a fixed period, T , the velocity of a point and the velocity after the T period of this point often show some similarity and consistency, which function together by repeatability of driving behavior and stationarity of road conditions. We assume that cyclical change of travel speed conforms to a common expression of Fourier series [1], a complete cycle expression in a unit time interval shown as

$$v_r(t) = A_0 + \sum_{k=1}^n A_k \sin(k\omega t + \varphi_k), \quad (1)$$

in which $A_k \sin(k\omega t + \varphi_k) = a_k \cos k\omega t + b_k \sin k\omega t$ is harmonic of n order. Consider the truth of travel speed prediction occurring from 0 a.m. to 12 p.m., where (1) has a general form:

$$v_r(t) = A_0 + \sum_{k=1}^n (a_k \cos k\omega t + b_k \sin k\omega t), \quad (2)$$

in which $\omega = 2\pi/T$, which is a frequency index defined as cycles per unit time. T is the cycle time. k is the index of cycle series; n represents the order of periodic elements in the series. a_k and b_k are parameters determined by historical data.

2.2. Autoregressive Integrated Moving Average Model. The autoregressive integrated moving average (ARIMA) model has always been used to solve stationarity, randomness, and periodicity in time series analysis and is one of the most general models used to predict spot speed by its past speed data. The periodicity of traffic speed has been considered in the above analysis, but we should remove the periodic trend in daily speed data before using the ARIMA model to forecast travel speed. The remaining part represents variations in real-time specific traffic conditions.

In fact, ARIMA is the combination of the two algorithms: AR and MA, with I representing the integrated term. A nonseasonal ARIMA model is classified as ARIMA (p, d, q) model, in which AR means autoregressive, p is the number of the autoregressive term, MA means moving average, q is the order of the moving average, and d is the number of difference when the time series is steady. The mathematical representation of an ARIMA (p, d, q) is as follows:

$$\varphi_p (1 - B)^d X_t = \theta_0 + \theta_q (B) \mu_t, \quad (3)$$

where X_t is an original data series; μ_t is a white noise sequence, which is a sequence of random variables where the mean is zero and variance is σ^2 ; B represent lag operators, $BX_t = X_{t-1}$; φ_p is autoregression operators, $\varphi_p(B) = (1 - \varphi_1 B - \dots - \varphi_p B^p)$, and p is autoregression order of model; θ_q is moving average operators, $\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$, and q is order of moving average; θ_0 is a parameter, $\theta_0 = v(1 - \varphi_1 - \varphi_2 - \dots - \varphi_p)$; and v is a mean value.

The process of ARIMA modeling and analyzing is as follows, which is implemented by **R**:

(1) Smoothing the data: if the data sequence is not stable, we should differentiate the data, and the times of difference are d , until we get a stationary time series.

(2) Model identification: the initial ARIMA (p, d, q) model gained from autocorrelation function (ACF) and partial autocorrelation function (PACF) of the processed time series could determine autoregression order (p) and order of moving average (q) preliminary.

(3) Parameter estimation and model diagnostic: when we gain the coefficients of the initial, we should test the significance of coefficients in the model and at the same time test the white noise of the model.

(4) Forecast and analyze the data using a model with appropriate parameters.

2.3. The Spatial Correlation Analysis. The road consists of connected sections; each spot on the road has spatial accessibility. We can learn from the theory of spatial analysis that the relationship between the difference of attribute values in space and the distance between two points obeys the first law of geography [13]. This means that, from a statistical point of view, the closer the distance between two points, the higher the degree of similarity they have. The spatial correlation of traffic flow would reduce with the increase of distance in a certain space. For a section of same spatial distance in a road, the spatial correlation generally increased, accompanied by the increase of traffic load [14].

Correlation index ρ of statistics could analyze the spatial correlation of travel speed (study on remote monitor system based on 3G mobile system). Assume that there are two points A and B in the adjacent position, x_{Ai} , x_{Bi} is the spot speed of A and B at the same time i , and the correlation index is as follows:

$$\rho_{AB} = \frac{\text{cov}(x_{Ai}, x_{Bi})}{\sqrt{D(x_{Ai})}\sqrt{D(x_{Bi})}}, \quad (4)$$

in which $\text{cov}(x_{Ai}, x_{Bi})$ is the covariation of speed A and B at the time i , $\text{cov}(x_{Ai}, x_{Bi}) = (1/n) \sum_{i=1}^n (x_{Ai} - \bar{x}_A)(x_{Bi} - \bar{x}_B)$; $D(x_{Ai})$ and $D(x_{Bi})$ are the time series variance of speed in A and B at time i .

As the vehicle has liquidity, the vehicle would arrive downstream from upstream after a certain period of time; the upstream traffic flow would be influenced if congestion or failure occurred in downstream. Therefore, we introduce lagged value while analyzing the spatial correlations. Assuming that a car traveled from A to E, with

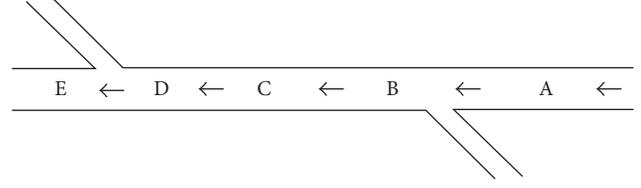


FIGURE 1: Location of 5 adjacent points in one road.

time-lagged value expressed as γ , the correlation index is changed as

$$\rho_{AB\gamma} = \frac{\text{cov}(x_{Ai}, x_{B(i-\gamma)})}{\sqrt{D(x_{Ai})}\sqrt{D(x_{B(i-\gamma)})}}. \quad (5)$$

Correlation response is the degree of linear correlation between different groups of data. After repeated verification, two points adjacent to the forecast point upstream and downstream are suitable for selection. If the forecast point has only one adjacent point, the sample size is not sufficient to predict accurately. If the predicted point has too many adjacent points, it will produce invalid calculations and reduce the computational efficiency. Assuming that there are five adjacent points on one road as shown in Figure 1, we can build a multiple linear regression (MLR) model to forecast travel speed of point C as follows:

$$\begin{aligned} x_C(t) = & \lambda_0 + \lambda_1 x_A(t - \gamma_{Amin}) \\ & + \lambda_2 x_A(t - (\gamma_{Amin} + 1)) + \dots \\ & + \lambda_i x_A(t - \gamma_{Amax}) + \lambda_{i+1} x_B(t - \gamma_{Bmin}) \\ & + \lambda_{i+2} x_B(t - (\gamma_{Bmin} + 1)) + \dots \\ & + \lambda_j x_B(t - \gamma_{Bmax}) + \lambda_{j+1} x_D(t - \gamma_{Dmin}) \\ & + \lambda_{j+2} x_D(t - (\gamma_{Dmin} + 1)) + \dots \\ & + \lambda_m x_D(t - \gamma_{Dmax}) + \lambda_{m+1} x_E(t - \gamma_{Emin}) \\ & + \lambda_{m+2} x_E(t - (\gamma_{Emin} + 1)) + \dots \\ & + \lambda_n x_E(t - \gamma_{Emax}) + \varepsilon, \end{aligned} \quad (6)$$

in which $\lambda_0, \lambda_1, \dots, \lambda_i, \lambda_{i+1}, \dots, \lambda_j, \lambda_{j+1}, \dots, \lambda_m, \lambda_{m+1}, \dots, \lambda_n$ are coefficients of MLR model. 1 to i is the number of lagged value between A and C, $i + 1$ to j is the number of lagged value between B and C, $j + 1$ to m is the number of lagged value between C and D, $m + 1$ to n is the number of lagged value between C and E. And n is the total amount of time-lagged value at points A, B, D, and E and C, which is an integer greater than or equal to zero. And i, j, m, n all come from formula (5). The value of λ is also obtained from a large amount of data by multiple linear regression from formula (5). $\gamma_{Amin}, \gamma_{Bmin}, \gamma_{Dmin}, \gamma_{Emin}, \gamma_{Amax}, \gamma_{Bmax}, \gamma_{Dmax}, \gamma_{Emax}$ are the minimum and maximum time-lagged value between point A and C, B and C, C and D, as well as C and E. The calculated time-lagged values come from formula

(5). The value of each point and the point C after the spatial correlation analysis can be determined, in which, we define the smallest γ of one point and point C as the minimum value and the largest one as the maximum value. The time-lagged values are continuous between the maximum value and the minimum value. $x_A(t - \gamma_{Amin}), \dots, x_E(t - \gamma_{Emax})$ denote the speed of points A, B, D, and E at the period of $t - \gamma_{Amin}, \dots, t - \gamma_{Emax}$, respectively. Note that we can determine the number of terms of multiple linear regression and the value of γ in each term only with specific data. ε is the random error.

For application in the actual engineering project, we should test statistical information at first, including model verifying and a goodness-of-fit test. Then, established time series or predicted value of time series can forecast the value of travel speed of the target position next period.

2.4. Hybrid Model. The travel speed has strong temporal and spatial correlation in the road network. The speed change in the next period can be regarded as a continued speed variation rule of what occurred in the last period; on the other hand, travel speed can be affected by traffic conditions in nearby links. An alternative way is to conduct the change of travel speed as a combination of cyclic and definitive components determined by real-time transformation, space-time relationship, specific road conditions, and so on.

In this study, we consider the time series and space influence of speed at the same time, which purposely increases the accuracy of short-time speed prediction. The reliability and accuracy of traffic speed prediction depend on a comprehensive understanding of every part of speed in the above analysis. Therefore, we assume that traffic speed is constituted of three components based on the discussion in the introduction: a periodic trend, a residual time series, and the space-time part. The structure of the proposed model is shown as follows:

$$\hat{y}_t = \sum_i^n w_{ti} y_{ti} + \varepsilon, \quad (7)$$

where \hat{y}_t is the predictor value of travel speed at time t , y_{ti} is predictor value of method i at time t , w_{ti} are corresponding parameters of y_{ti} , of which values can be obtained through the regression of y_{ti} some time ago; ε is a constant.

Calculating this formula usually requires taking advantage of analysis software and then selecting the most suitable model according to the specific conditions.

3. Case Study

The speed data used in the study were collected on a northeastward segment of I-394 freeway stretching between Trunk Highway 100 and Egret Blvd in Twin Cities Metro area, Minnesota [15], which is a two-way six-lane section. This freeway segment experiences significant speed decline during afternoon peak hours. Five adjacent stations are located on the selected segment to collect traffic speed data, in which point C is the one we try to predict. The distance between adjacent stations is about one mile. In order to ensure traffic stability on the road, we specifically chose points far from the import and export ramps that have large amounts of traffic,



FIGURE 2: Location of 5 adjacent stations on I-394 segment, from station S979 to station S983.

and we avoided the intersection to reduce the impact of traffic lights on the speed deliberately. The location of the segments and stations is shown in Figure 2.

The travel speed data were collected once every 5 minutes, 24 hours a day, from July 14, 2015, to July 13, 2016. The data has a long time span, so we removed the abnormal value caused by extreme weather conditions, traffic flow breakdown, and so on. The missing data for the 5 stations are less than 5%, and, in order to ensure the integrality of selected speed data for model validation in this study, we adopt historical average data to replenish missing data and bad data.

3.1. Periodic Analysis by Fourier Series. It is known that the traffic speed pattern on weekdays is quite different from that on weekends; traffic flow during weekends is normally smooth. Thus, this study chose the data on weekdays to express periodicity of speed change. Taking Monday as an example, Figure 3(a) exhibits the law of speed variation and Figure 3(b) exhibits historical median travel speed in different days integrally. It is observed that there is one set of peak hours (from 3:00 p.m. to 7:30 p.m.), and the peak values of different weekdays are variable.

Through the analysis shown in Figure 3, we establish a two-dimensional Fourier series to match the periodicity of speed changed by day. For example, we plot the speed of every Thursday of 53 weeks and fit the scattered points, as shown in Figure 4.

General model of Fourier series after expansion in 8 stages is given by

$$\begin{aligned} f(x) = & a_0 + a_1 \cos(x\omega) + b_1 \sin(x\omega) + a_2 \cos(2x\omega) \\ & + b_2 \sin(2x\omega) + a_3 \cos(3x\omega) + b_3 \sin(3x\omega) \\ & + a_4 \cos(4x\omega) + b_4 \sin(4x\omega) + a_5 \cos(5x\omega) \\ & + b_5 \sin(5x\omega) + a_6 \cos(6x\omega) + b_6 \sin(6x\omega) \\ & + a_7 \cos(6x\omega) + b_7 \sin(6x\omega) + a_8 \cos(7x\omega) \\ & + b_8 \sin(7x\omega). \end{aligned} \quad (8)$$

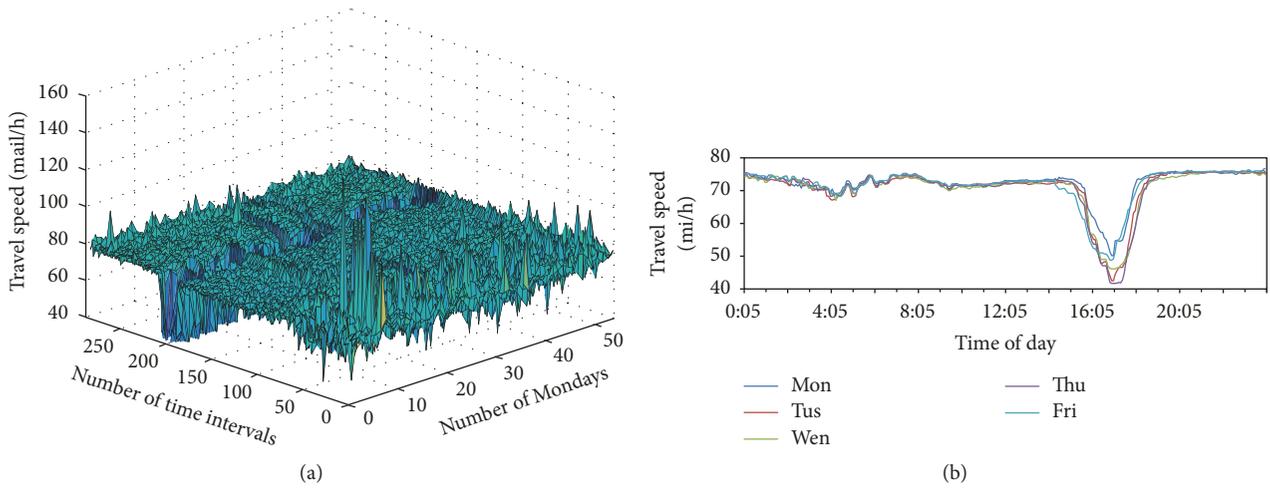


FIGURE 3: (a) 53 Monday speed change. (b) Historical median travel speed change from Monday to Friday.

TABLE 1: Coefficient of Fourier series.

	Monday	Tuesday	Wednesday	Thursday	Friday
a_0	71.64	70.31	70.22	70.2	71.11
a_1	0.4058	-0.3655	0.1058	-0.3338	0.7677
b_1	2.246	4.152	4.325	5.006	3.813
a_2	3.314	4.905	4.877	5.242	4.82
b_2	0.3869	1.684	1.351	2.028	-0.1861
a_3	2.673	4.239	3.622	4.623	1.957
b_3	-2.563	-2.571	-3.27	-3.035	-3.398
a_4	-1.992	-1.878	-2.676	-2.289	-3.225
b_4	-1.133	-2.757	-1.872	-3.035	-0.3212
a_5	-2.093	-3.453	-2.8	-3.518	-1.455
b_5	1.293	0.1584	1.268	0.4451	1.89
a_6	0.5955	-0.5626	0.6319	-0.2943	1.421
b_6	1.801	2.252	1.963	2.234	0.895
a_7	1.324	1.688	1.463	1.391	0.826
b_7	-0.4243	0.4932	-0.3302	0.4009	-0.9971
a_8	0.1335	1.055	0.3197	0.9583	-0.3754
b_8	-1.174	-0.8825	-1.067	-0.868	-0.7175
ω	0.02423	0.02461	0.02415	0.02448	0.02412

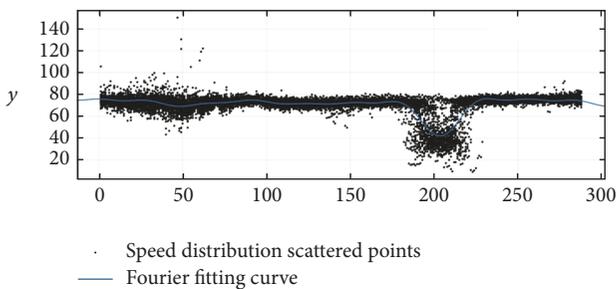


FIGURE 4: Speed distribution scattered points and Fourier fitting curve.

Using data from the past year, we can get the coefficient of Fourier series shown in Table 1.

The goodness-of-fit test results of each working day (as shown in Table 2) show that Fourier series cannot be a good fit. Values of R -square, which is considered one of the main indicators of the accuracy of the evaluation, are all less than 0.6, which means the accuracy of this fitting model is limited. Therefore, just using periodic trend analysis to forecast speed is not enough.

3.2. Process of ARIMA Modeling. We take the speed data of June 14th as an example. As shown in Figure 5, the residual part of the travel speed is the remaining of the trend part removed from the original data, which still shows some patterns. Dealing with volatile short-time prediction, the ARIMA model had been widely used, and its effectiveness and accuracy have been confirmed by a large number of experiments.

TABLE 2: Goodness-of-fit test.

	Monday	Tuesday	Wednesday	Thursday	Friday
SSE	$6.361e + 05$	$6.963e + 05$	$8.82e + 05$	$6.765e + 05$	$7.805e + 05$
R-square	0.3553	0.5079	0.4383	0.5627	0.4085
Adjusted R-square	0.3546	0.5074	0.4376	0.5622	0.4079
RMSE	6.399	6.758	7.606	6.661	7.155

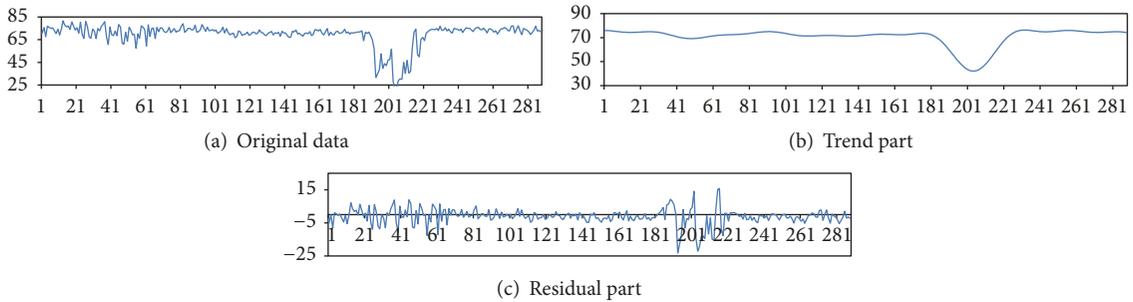


FIGURE 5: Dividing the speed into a trend and a residual part.

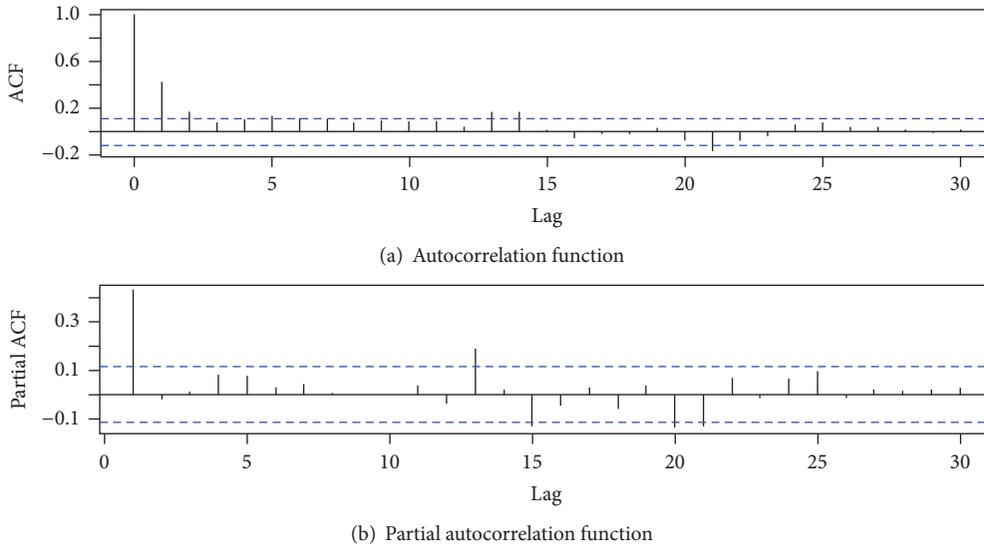


FIGURE 6: Plot of ACF and PACF in 14 June of Section C.

Through the analysis above focusing on the periodicity of speed data, we do not need the difference proposed to remove the data cycle in the process of ARIMA modeling, which means $d = 0$ in ARIMA model. Accordingly, ARIMA model translates into ARMA (p, q) model. The autocorrelation function and partial autocorrelation function of the residual part could inform us of the value of p and q .

From Figure 6, we know that there are couples of order beyond the confidence interval, which lead to a number of alternative models. It is usual to determine the specific parameters of the model by Akaike's information criterion (AIC). According to this criterion, value size is used as a primary basis for determining which model is the best. It is obvious that we could improve the goodness of fit by increasing the number of free parameters. AIC could express

the excellent properties of data fitting and avoid overfitting at the same time:

$$AIC = -2 \ln(L) + 2k. \tag{9}$$

In our case analysis, the most suitable model is ARIMA $(1, 0, 2)$. Through this model, we can obtain the speed residual part of the predicted value, and the confidence interval is 95%. The residual part of the prediction results is shown in Figure 7.

3.3. Discussion of Spatial Correlation. We can know the location of the detection point distribution from Figure 2, taking point C as the target location, and the sample cross correlation function (CCF) between point C and other points in different lags. In statistical studies, there is generally a strong correlation when the correlation coefficient is larger than

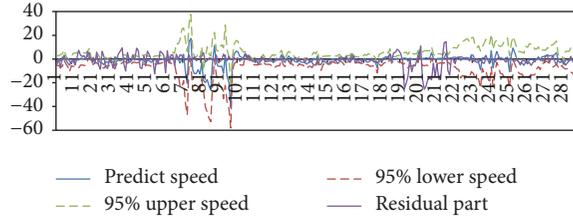


FIGURE 7: The residual prediction results.

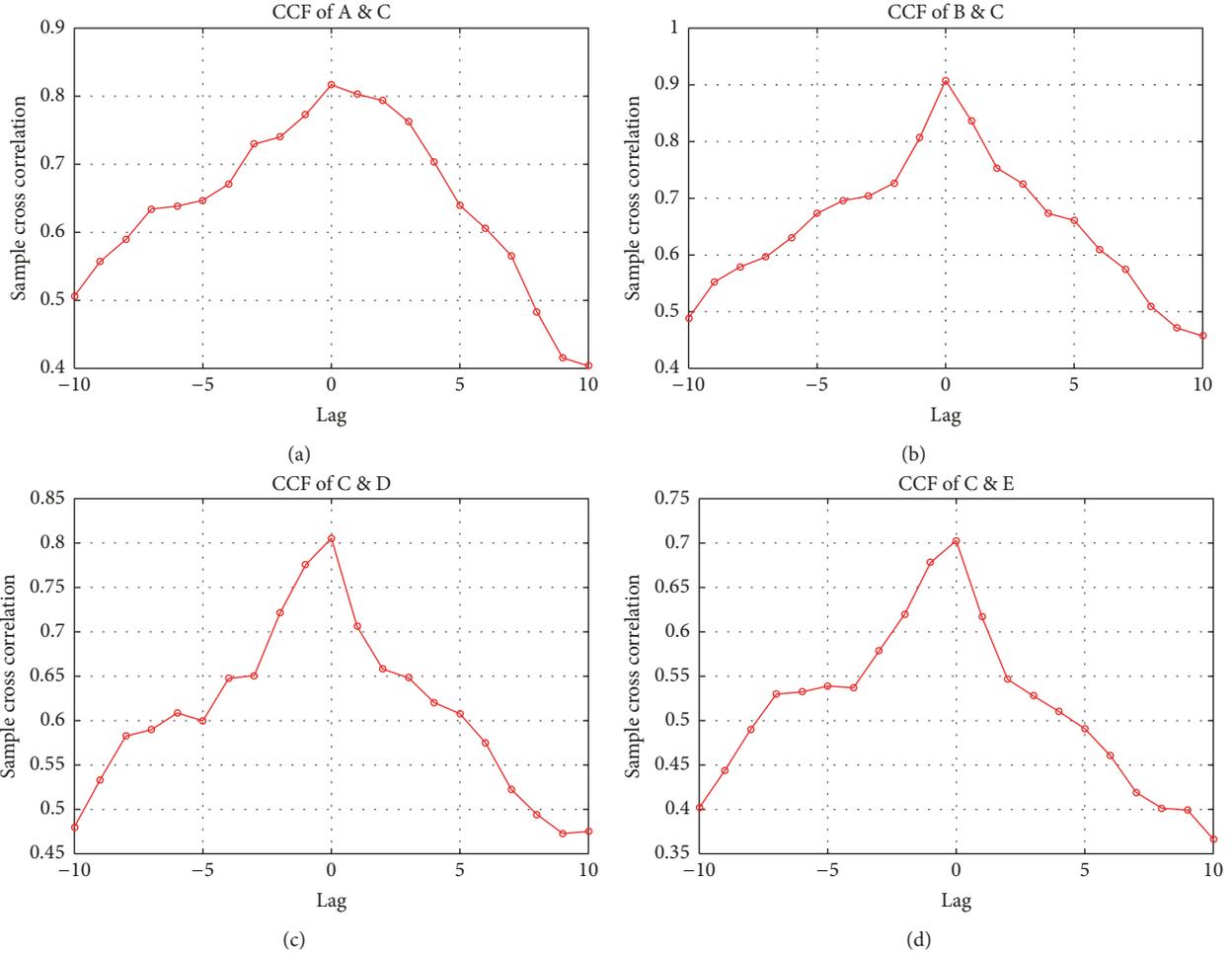


FIGURE 8: (a) CCF of speed between point A and point C. (b) CCF of speed between point B and Point C. (c) CCF of speed between point C and point D. (d) CCF of speed between point C and point E.

0.8. As shown in Figure 8, CCF gradually increases with the decrease of the lags' absolute value. Influence of upstream points is larger than that of downstream points on speed.

In our case, the CCF of E and C is less than 0.8; the relationship between these two points is ignored in the process of calculation. Therefore we consider that there are strong correlations between C and A, B, D, and that the largest lag is 1. The general expression of MLR model in point C is given by

$$\begin{aligned}
 x_c(t) = & \lambda_0 + \lambda_1 x_A(t) + \lambda_2 x_A(t-1) \\
 & + \lambda_3 x_B(t+1) + \lambda_4 x_B(t) + \lambda_5 x_B(t-1) \\
 & + \lambda_6 x_D(t).
 \end{aligned} \tag{10}$$

Parameters in the model can be obtained by linear fitting. The speed of point C can be forecasted according to real-time speed data and the model above. The predicted results show that the R -square is 0.8745 and the probability of the corresponding F statistic is close to 0. So, this MLR model is established.

This method can obtain higher accuracy of prediction results but will be easily affected by upstream and downstream traffic flow. If the upstream or downstream speed record failed, the abnormal value of prediction would appear.

3.4. Hybrid Model. In the above study, we get the result of prediction of time and space, and the final result was shown

TABLE 3: Evaluation index of different prediction method.

	RMAE	MAE	MAPE	mAPE
ARIMA	15.47	8.51	16.14%	93.56%
Spatial correlation Regression	6.63	4.76	6.61%	71.70%
Trend and ARIMA	7.50	4.90	8.61%	81.57%
Hybrid model	5.91	4.07	6.42%	64.18%

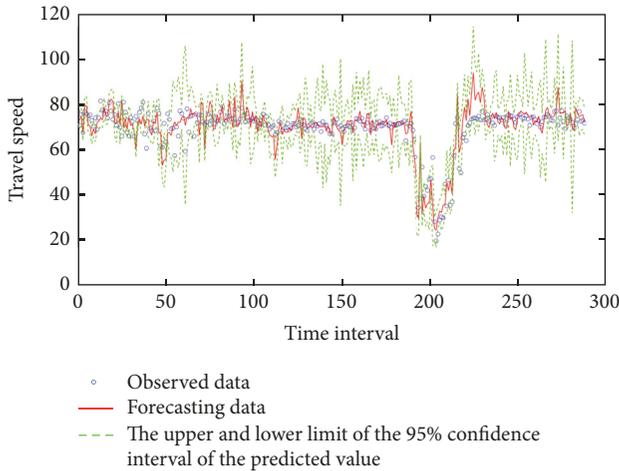


FIGURE 9: Results of hybrid model prediction.

in Figure 9. In order to improve the accuracy of the forecast, this case uses the data of every Tuesday in the past year as training data and the data of the previous hour of the forecast target as the testing data. All data are separated by 5 min.

In this case, there is no correlation between speed prediction time and space after being calculated, so we use a nonlinear regression model to estimate the parameters of the hybrid model. The red solid line stands for the results of predicted travel speed, the blue scatter represents the real speed data, and the green dash lines express the upper and lower bounds of the 95% prediction intervals. The model we proposed is able to predict traffic flow with a small degree of error in most situations. Generally speaking, the possibility of speed anomalies becomes greater where confidence interval amplitude becomes greater so that it can cover the abnormal data. The velocity prediction results show a large fluctuation between 210 and 250, and there is a certain error with the actual velocity. Similarly, there is such a phenomenon between 50 and 100. This is because the ARIMA model is still a kind of time series model, and, with the use of time series prediction methods, the forecast results will be subject to a period of time before the speed of the impact. This is also reflected in ACF and PACF when making ARIMA predictions.

4. Discussion

To objectively evaluate the effectiveness of different prediction models, four types of indexes are introduced for comparison: the root mean squared error (RMSE), the mean absolute

error (MSE), the mean absolute percentage error (MAPE), and the maximum absolute percentage error (mAPE), with the results shown in Table 3.

As seen from Table 3, the predicted result of single time series method is the worst of these four methods. Spatial correlation predicted is better with experimental data, but the model assumes that the points on road are close to adjacent and that the road alignment is stable with no mutation. When the cross correlation is less than 0.8, the correlation does not exist. The method combining periodicity trend and residual had higher accuracy but did not consider emergent events happening in upstream and downstream. Compared with those above methods, the indexes of the hybrid model have the top performance. Generally, the hybrid model prediction based on spatial-temporal fusion is better than single time series prediction and spatial regression estimation prediction. Availability of this method is verified with a higher forecasting precision.

The results of the periodic analysis show that the points in the case are discrete at a speed between 0 a.m. and 7 a.m. and at 3 p.m. to 6 p.m. will appear to decrease speed and then return to normal again. These rules exist every Monday. Therefore, the decision-makers can set the maximum and minimum speed at 0 a.m. to 7 a.m., in order to stabilize the speed difference between vehicles, improving traffic safety. ARIMA prediction residuals are basically in accordance with actual residuals, but it is important to note that ARIMA predictions are affected by data changes, so there is some lag in the prediction of actual changes, but the results are basically within the 95% confidence interval. The results of spatial correlation analysis show that there is a strong correlation between A, B, D and C; that is, if there are emergencies at points A, B, and D, warnings should be set at point C in time to reduce congestion diffusion.

5. Conclusion

This paper proposed a hybrid model of multistep short-time travel speed prediction with spatial and temporal fusion. This model divides travel speed into three parts: a periodic trend, residual part, and the mutation influenced by upstream and downstream traffic flow. The change of the periodic part is generally by week, and we can gain the periodic trend expression by using Fourier series fitting historical data. Residual part is the difference between real speed and the value of time corresponding to the periodic function. The residual part is then further analyzed by the ARIMA model, the purpose of which is to forecast the change in a short time. The aim of the spatial analysis regression model is to achieve the influence of traffic incidents. Therefore, the model

can improve the accuracy of the prediction of the speed and provide feedback on the law of speed at the same time, especially the periodic variation. It provides the theoretical basis for improving the service quality of ITS and traffic information systems. Freeway travel speed data was collected from five sensors located at I-394 in Minnesota. Through the calculation of multistep, we gained the predicting result of the proposed model. Comparisons between the proposed hybrid model with others indicated that treating the time and space comprehensively can generate better forecasting accuracy and reliability in terms of RMSE, MSE, MAPE, and mAPE.

Based on this study, there are two conclusions worth noting. First, there is no significant correlation between the speed predicted by time series and spatial correlation analysis. Therefore, we can only use a nonlinear function to fit data when taking calculations; concrete formulas need to be calculated repeatedly in SPSS or MATLAB according to the actual data, which means the rate is low in initial setting. Furthermore, the function needs reeducation after a period of time, to prove there is no change of the expression. Secondly, travel speed has always been considered to exhibit strong day-to-day cyclical patterns. But we found the speed changes and extremum of different days in a week are different by comparing with real data. Therefore, the periodic law can be expressed more accurately on a week-to-week basis.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research is supported by Fundamental Research Funds for the Central Universities (CXZZ13_0118), the Scientific Research Foundation of Graduate School of Southeast University (YBJJ1457), the Key Project of National Natural Science Foundation of China (51338003), the National Natural Science Foundation of China (5151101143), Projects of International Cooperation and Exchange of the National Natural Science Foundation of China (5151101143), and Jiangsu Planned Projects for Postdoctoral Research Funds (1701082B).

References

- [1] D. S. Dendrinos, "Traffic-flow dynamics: A search for chaos," *Chaos, Solitons & Fractals*, vol. 4, no. 4, pp. 605–617, 1994.
- [2] X. Fei, C. C. Lu, and K. Liu, "A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1306–1318, 2011.
- [3] L. Liu, EMD based hybrid models in short term traffic speed forecast, 2014.
- [4] Y. Zhang, Y. Zhang, and A. Haghani, "A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model," *Transportation Research Part C: Emerging Technologies*, vol. 43, no. 1, pp. 65–78, 2013.
- [5] 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.
- [6] E. R. Ziegel, "Forecasting and Time Series: An Applied Approach," *Technometrics*, vol. 36, no. 4, pp. 434–434, 1994.
- [7] R. Lund, "Time Series Analysis and Its Applications: With R Examples," *Journal of the American Statistical Association*, vol. 102, no. 479, pp. 1079–1079, 2007.
- [8] M. G. Karlaftis and E. I. Vlahogianni, "Memory properties and fractional integration in transportation time-series," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 4, pp. 444–453, 2009.
- [9] J. C. van Lint, "Reliable real-time framework for short-term freeway travel time prediction," *Journal of Transportation Engineering*, vol. 132, no. 12, pp. 921–932, 2006.
- [10] T. L. Pan, A. Sumalee, R. X. Zhong, and N. Indra-Payoong, "Short-term traffic state prediction based on temporal-spatial correlation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1242–1254, 2013.
- [11] Y. Zou, X. Hua, Y. Zhang, and Y. Wang, "Hybrid short-term freeway speed prediction methods based on periodic analysis," *Canadian Journal of Civil Engineering*, vol. 42, no. 8, pp. 570–582, 2015.
- [12] J. J. Tang, F. Liu, Y. J. Zou, W. B. Zhang, and Y. H. Wang, "An improved fuzzy neural network for traffic speed prediction considering periodic characteristic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, pp. 2340–2350, 2017.
- [13] Y. C. Chen, T. Cao, C. Chen et al., "Molecular bandgap engineering of bottom-up synthesized graphene nanoribbon heterojunctions," *Nature Nanotechnology*, vol. 10, no. 2, pp. 156–160, 2015.
- [14] T. Zhao, Y. Zhang, Y. Zhou, and S. Feng, "Spatial cross correlations of traffic flows on urban road networks," *Qinghua Daxue Xuebao/Journal of Tsinghua University*, vol. 51, no. 3, pp. 313–317, 2011.
- [15] The Minnesota Department of Transportation collects traffic data on the freeway system throughout the Twin Cities Metro area, <http://data.dot.state.mn.us/datatools>.



Hindawi

Submit your manuscripts at
www.hindawi.com

