View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by HKU Scholars Hub

The HKU Scholars Hub The University of Hong Kong 香港大學學術庫

Title	Short-term traffic speed forecasting based on data recorded at irregular intervals
Author(s)	Ye, Q; Szeto, WY; Wong, SC
Citation	IEEE Transactions on Intelligent Transportation Systems, 2012, v. 13 n. 4, p. 1727-1737
Issued Date	2012
URL	http://hdl.handle.net/10722/152648
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Short-Term Traffic Speed Forecasting Based on Data Recorded at Irregular Intervals

Qing Ye, W. Y. Szeto, and S. C. Wong

***Abstract*—Recent growth in demand for proactive real-time transportation management systems has led to major advances in short-time traffic forecasting methods. Recent studies have introduced time series theory, neural networks, and genetic algorithms to short-term traffic forecasting to make forecasts more reliable, efficient, and accurate. However, most of these methods can only deal with data recorded at regular time intervals, which restricts the range of data collection tools to presence-type detectors or other equipment that generates regular data. The study reported here is an attempt to extend several existing time series forecasting methods to accommodate data recorded at irregular time intervals, which would allow transportation management systems to obtain predicted traffic speeds from intermittent data sources such as Global Positioning System (GPS). To improve forecasting performance, acceleration information was introduced, and information from segments adjacent to the current forecasting segment was adopted. The study tested several methods using GPS data from 480 Hong Kong taxis. The results show that the best performance in terms of mean absolute relative error is obtained by using a neural network model that aggregates speed information and acceleration information from the current forecasting segment and adjacent segments.**

***Index Terms*—Autoregressive integrated moving average (ARIMA), combined forecasting, exponential smoothing method, Holt's method, irregularly spaced time series data, neural network, short-term traffic speed forecasting.**

I. Introduction

IRREGULARLY spaced or intermittent data is a problem commonly encountered in many research fields. In this paper, irregularly spaced data refer to information uploaded to data collectors at unequal time intervals. Figs. 1 and 2 show the difference between irregularly spaced data and classic regularly spaced data. Forecasting based on irregularly spaced data is usually more complicated and less accurate than that based on regularly spaced data. One reason for this is the lack of information caused by the irregularity of the time intervals. Regularly spaced data has a constant time interval, and it is reasonable and simple to use this constant interval as a forecasting step. However, with irregularly spaced data, the time intervals are variable. The most reasonable forecasting step for irregularly spaced data is the smallest time interval, yet this still results in missing information, particularly when the time interval is very large.

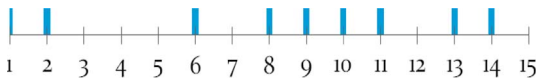

Fig. 1. Irregularly spaced data.

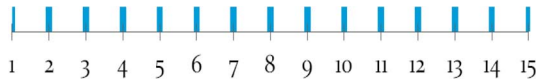

Fig. 2. Regularly spaced data.

Many researchers have extended the classic time series methods in an attempt to deal with the problem of irregularly spaced data. Croston [1] developed a method for providing more accurate estimates of stock control. This method separately applies exponential smoothing to the intervals between nonzero demands and the magnitude of these demands by treating them as a continuous time series according to their indices. Jones [2] later attempted to apply the maximum-likelihood fitting of autoregressive moving average (ARMA) models to time series with missing observations. However, this method is useful only when the number of missing observations is very small. Wright [3] further developed the simple exponential smoothing method and Holt's method by extending the use of weighted averages for irregularly spaced data. In 1995, Cipra *et al.* [4] expanded on Wright's modification of Holt's method by proposing an extended version of the Holt–Winters method for analyzing data taken at irregular time intervals. In 2004, Carmo and Rodrigues [5] compared the methods of Wright and Croston using a neural network method to show how a neural network can be used for filtering and forecasts. Willemain *et al.* [6] suggested a bootstrap method for intermittent data in service part inventories. Velicer and Colby [7] compared four missing-data solutions (deletion, mean substitution, mean of adjacent observations, and maximum-likelihood estimation methods) commonly used with autoregressive integrated moving average (ARIMA) models for time series under 50 different data-missing conditions and showed that the maximum-likelihood estimation method performed equally well under all of the conditions tested. In 2006, Cipra [8] modified the double smoothing method in a manner similar to that used by Wright to alter Holt's method and suggested that the modified double smoothing method performs better than the modified Holt

Manuscript received October 21, 2011; revised February 20, 2012 and May 21, 2012; accepted May 29, 2012. Date of publication June 21, 2012; date of current version November 27, 2012. This work was jointly supported by a grant from the Research Grants Council of HKSAR, China, under Project HKU7176/07E, Grant 200902172003 from the Hui Oi Chow Trust Fund, Grant 201011159026 and Grant 201111159056 from the University Research Committee of the University of Hong Kong, a Research Postgraduate Studentship from the University of Hong Kong, and an Outstanding Researcher Award from the University of Hong Kong. The Associate Editor for this paper was W.-H. Lin.

The authors are with the Department of Civil Engineering, University of Hong Kong, Pokfulam, Hong Kong (e-mail: yeq09@graduate.hku.hk; ceszeto@hku.hk; hhecwsc@hku.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2012.2203122

1524-9050/$31.00 © 2012 IEEE

method. However, this comparison may not be accurate, because Cipra did not optimize the coefficients in this study. Altay *et al.* [9] compared a modification of Croston's method and Wright's modified version of Holt's method in 2008 and found that using Croston's method to forecast demand resulted in the minimization of inventory levels, whereas using the modified Holt method resulted in superior customer service. Cipra and Hanzak [10] devised a method involving exponential smoothing to order m for irregular time series.

Another major research effort focused on the use of time series to forecast traffic state and travel time. In 1997, Smith and Demetsky [11] compared four methods, i.e., the time series method, the historical average method, neural networks, and nonparametric regression models, to forecast traffic flow 15 min ahead. Results indicated that the nonparametric regression model was the most accurate. However, Smith *et al.* [12] reached a different conclusion, showing that the seasonal ARIMA (SARIMA) model delivers results that are statistically superior to those of nonparametric models. William and Hoel [13] proposed the SARIMA model as a theoretical basis for modeling traffic data. Szeto *et al.* [14] developed an integrated method incorporating both the SARIMA model and the cell transmission model. Forecasting results suggest that these methods are effective for predicting real-time traffic flow data. Tan *et al.* [15] suggested an aggregation method in which the moving average (MA) and exponential smoothing (ES) methods and the ARIMA model are used to forecast three relevant time series and the predicted results are aggregated in a neural network model to generate the final forecast. Simroth and Zähle [16] presented a new algorithm for predicting the remaining travel times of long-range trips. The algorithm made use of nonparametric distribution-free regression models and could reduce the average relative error by half when comparing with conventional methods. Sun and Xu [17] proposed a novel variational inference approach to infinite mixtures of Gaussian processes using the Dirichlet process prior and applied it to traffic flow prediction. This approach was shown to outperform the Bayesian network approach, the random walk approach, and ridge regression. Khosravi *et al.* [18] introduced the delta and Bayesian techniques for the construction of prediction intervals of travel time. The results indicated that the delta technique outperformed the Bayesian technique in terms of narrowness of prediction intervals with satisfactory coverage probability, but the Bayesian technique outperformed the delta technique in terms of robustness against the neural network structure and coverage probability. Wang *et al.* [19] applied traffic state prediction models for real-time freeway network traffic surveillance, and the results were satisfactory. Soriguera and Robusté [20] presented a two-level data fusion methodology for short-term travel time prediction, where the first level fused the instantaneous travel times estimated from different types of loop detectors and the second level combined the fused travel times obtained in the first level with the measured travel times obtained the toll ticket data. Results showed that the overall fused travel times are more accurate than the instantaneous travel times on average. Chan *et al.* [21] proposed the hybrid exponential smoothing and Levenberg–Marquardt algorithm to train neural network models for short-term traffic flow forecasting. Exponential smoothing was used to preprocess traffic flow data by removing the lumpiness from collected traffic flow data, before training the weights. Results showed that the neural network model developed using the proposed approach outperformed those neural network models without preprocessing, with MA for preprocessing, and with weighted MA for preprocessing. Tchrakian *et al.* [22] adopted traffic flow forecasting method based on spectral analysis. The errors were found to be comparable with those based on time series and neural network models.

The aforementioned studies have together made significant progress in forecasting traffic state by treating traffic condition data as time series data. However, most of the existing methods can only deal with regularly spaced time series data, because they involve the use of presence-type detectors to obtain the necessary data. Given that it is impractical to install presence-type detectors at all links in a city, some alternative ways of generating information on traffic conditions have been developed, such as floating cars that upload GPS data. However, the data collected using these methods are irregularly spaced time series data, which are not suitable for most of the existing time series traffic flow forecasting methods. To the best of our knowledge, only Chen *et al.* [23] and van Lint *et al.* [24] proposed hybrid neural networks for travel flow and travel time prediction under missing data conditions.

The study reported here is an attempt to propose an alternative method to address the problem of using intermittent data for forecasting traffic conditions. The study expands on the classic methods for dealing with intermittent time series data to allow traffic conditions to be forecast using irregularly spaced data. It endeavors to compensate for the loss of forecasting accuracy due to missing information (compared to the situation of having data for every second) by introducing acceleration information and information from adjacent points. Our contribution is in proposing a speed-forecasting method that incorporates the acceleration information and can make more accurate predictions than some of the existing methods.

II. Data

The study used the GPS data of 480 Hong Kong taxis collected over 14 weeks from June 29, 2009, to October 4, 2009. The GPS equipment installed in the taxis uploaded their coordinates, speeds, driving directions, and the uploading time to an information center approximately once every 30 s. To eliminate bias from taxi drivers intentionally driving slowly to look for passengers, a flag-on attribute (where a taxi for hire $= 0$ and a taxi not for hire $= 1$) was also uploaded. Only GPS data with a flag-on equal to 1 were used for forecasting, because taxis with this attribute were assumed to behave more like normal vehicles.

The study used a GIS technique to locate the taxis' position according to their coordinates on a road map. A 2-km-long section of Princess Margaret Road approaching the Cross Harbour Tunnel in Hong Kong was selected. Cross Harbour Tunnel is the busiest tunnel connecting Hong Kong Island and Kowloon, and is one of the four busiest tunnels in the world. More than 120 000 vehicles go through the tunnel everyday.

TABLE I
CHARACTERISTICS OF THE DATA SET

Data set	Number of records	Speed (km/h)			
		Maximum	Minimum	Mean	Standard deviation
Calibration set	181,952	120	0	22.85	25.38
Validation set	33,509	95	0	20.58	24.26

TABLE II
DISTRIBUTION OF INTERVAL LENGTHS

Interval (seconds)	Percentage of observations
0-400	60%
401-800	13%
801-1200	8%
1201-1600	5%
1601-2000	3%
2001-2400	2%
2401-33600	9%

GPS DATA

Dt	Tm	SpeedKmHr	Segment
2009-6-28	84247	52.522720	8
2009-6-28	84638	77.709920	7
2009-6-28	90813	69.653720	6
2009-6-28	91344	61.301200	9
2009-6-28	91345	70.264880	8
2009-6-28	91921	52.948680	7
2009-6-28	92605	56.152640	6
2009-6-28	94230	56.967520	6
2009-6-28	94452	68.764760	5
2009-6-28	95442	90.488720	7
...	...	...	...

Fig. 3. GPS data sample.

The road section we chose was the main road leading to the Cross Harbour Tunnel. There are always traffic jams between 7:00 a.m. and 9:00 p.m. on weekdays. There are limited flyovers and ramps along this road section, and limited merges and diverges. There are no signalized intersections along this road section, and the posted speed limit is 70 km/h.

The studied section was divided into twenty 100-m segments to generate 20 forecasting points. Data for the first 12 weeks were used for calibration, and data for the last two weeks were used for validation. Table I shows the characteristics of speed collected in each data set. The average time interval between two observations is 841.21 s with a standard deviation of 1847.41 s. Table II shows the distribution of interval lengths between two consecutive observations. As shown in this table, the percentage of observations decreases as the length increases, and more than 90 percent of the observations fall in the range with the length less than 2400 s.

After data processing, the GPS data contained four items: 1) date (Dt); 2) time from the start of recording in seconds (Tm); 3) spot speed in kilometers per hour (SpeedKmHr); and 4) segment number (segment), as shown in Fig. 3. From the second and final columns in this figure, we can see that the observations for Segment 7 are irregularly spaced.

III. METHODS

The study adopted several methods to make speed forecasts based on irregularly spaced data and attempted to improve the forecasting results using acceleration information and correlations between data segments. The methods used were the naive method, Wright's modification of the simple exponential smoothing method, and Wright's modification of Holt's method [3]. The ARIMA model was used to forecast the acceleration information. Neural networks were used to aggregate the results from various methods. We also incorporated the acceleration information and information from adjacent segments to improve the forecasting performance. For comparison, an extrapolation method using the latest observed and predicted acceleration information was adopted. A neural network without acceleration information was also used for comparison.

A. Existing Methods

1) Naïve Method: With the naive method, the latest observation is used as the forecast for the next and subsequent periods. This serves as the worst-case scenario and can be considered a reference method, against which more complex methods can be compared. It can be mathematically expressed as follows:

$$\hat{y}_{t+k} = y_t \tag{1}$$

where y_t is the speed observed at time t, and $\hat{y}_{t+k}$ is the forecast made at time $t+k$. k is the difference between the time to predict the next speed and the time with the latest speed observation.

2) Simple Exponential Smoothing Method With Wright's Modification: This method represents a compromise between two extreme prediction methods: 1) the naive method, in which all of the weight is placed on the latest observation, and 2) the MA method, in which all of the observations are weighted equally to obtain the average. The exponential smoothing method gives the most weight to the latest observation and gives exponentially decreasing weight to more distant observations. All of the exponential smoothing methods have a recursive form as the basic principle

$$\begin{aligned}\text{new estimate} = (1 - \text{parameter}) \times \text{previous estimate} \\ +\text{parameter} \times \text{new observation}.\end{aligned} \tag{2}$$

The simple exponential smoothing method is the most basic method to follow this recursive form. The mathematical expression of the simple exponential smoothing method is

$$\begin{aligned}\hat{y}_{t+1} = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} \\ +\alpha(1-\alpha)^3 y_{t-3} + \cdots\end{aligned} \tag{3}$$

where α is a coefficient ranging from 0 to 1.

The recursive form of this method is

$$\hat{y}_{t+1} = (1-\alpha)\hat{y}_t + \alpha y_t. \tag{4}$$

However, this form cannot be easily applied to data sets with irregular time spacing. Wright applied the method to irregularly spaced data by treating the exponentially smoothing

method as an exponentially weighting process and eliminating the weights corresponding to the missing data. According to Wright's modification, the extended form of this method for irregularly spaced data is

$$\hat{y}_n = A_n V_n \tag{5}$$

where A_n and V_n are the weighted average and the normalizing factor, respectively, which are defined as follows:

$$A_n = \sum_{i=-\infty}^{n} (1-\alpha)^{t_n - t_i} y_{t_i} \tag{6}$$

$$\frac{1}{V_n} = \sum_{i=-\infty}^{n} (1-\alpha)^{t_n - t_i}. \tag{7}$$

The recursive form of Wright's modification is then given by

$$\hat{y}_{t_n+1} = (1-V_n)\hat{y}_{t_n} + V_n y_{t_n} \tag{2}$$

$$V_n = \frac{V_{n-1}}{(b_n + V_{n-1})} \tag{9}$$

$$b_n = (1-\alpha)^{t_n - t_{n-1}} \tag{10}$$

$$\hat{y}_{t_n+k} = \hat{y}_{t_n+1}. \tag{11}$$

The aforementioned equations can be used to make forecasts from irregularly spaced data. In particular, the forecasts for times $t_n + 1, \ldots, t_n + k$ are all the same and based on the most recent observation y_{t_n}, the most recent forecast $\hat{y}_{t_n}$, the normalized constant V_n, and b_n according to (8)–(11).

3) Wright's Modification of Holt's Method: Holt's method is designed to deal with time series trends. It uses observations to provide estimates of both the level and slope of a time series. The recursive form of this algorithm is given by

$$L_t = \alpha y_t + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{12}$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \tag{13}$$

where L_t and T_t are the estimates of level and slope at time t, respectively. α and β are coefficients ranging from 0 to 1. The forecast for time $t + k$ at time t is

$$\hat{y}_{t+k} = L_t + kT_t. \tag{14}$$

One of the advantages of Holt's method is its flexibility, which is due to its use of two parameters. However, again, the aforementioned form cannot be applied to a data set with irregular time spacing. Wright's modification is thus applied to Holt's method

$$\hat{y}_{t_n+k} = L_n + kM_n \tag{15}$$

$$L_n = (1-V_n)\left(L_{n-1} + (t_n - t_{n-1})M_{n-1}\right) + V_n y_{t_n} \tag{16}$$

$$M_n = (1-U_n)M_{n-1} + U_n(L_n - L_{n-1})/(t_n - t_{n-1}) \tag{17}$$

where V_n and b_n are defined by (9) and (10), respectively, and U_n and d_n are their counterparts

$$U_n = \frac{U_{n-1}}{(d_n + U_{n-1})} \tag{18}$$

$$d_n = (1-\beta)^{t_n - t_{n-1}}. \tag{19}$$

In the recursive form, Wright's modification of Holt's method is simply derived by substituting α with V_n and β with U_n. The modified version of Holt's method still has two parameters and thus maintains its flexibility.

4) Forecasting Acceleration Information With the ARIMA Model: The ARIMA model was used to forecast the acceleration, so that this information could be incorporated into speed forecasting. Box and Jenkins introduced the ARIMA model and developed a methodology for data fitting [25].

The ARIMA model of order (p, d, q) can be written using the back-shift operator notation

$$\varphi(B)(1-B)^d(X_t - \mu) = \theta(B)\alpha_t \tag{20}$$

where

$$\varphi(B) = 1 - \varphi_1 B - \cdots - \varphi_p B^p \tag{21}$$

$$\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q \tag{22}$$

in which X_t is the value observed at time t; μ is the mean of all observed values; φ_i is an autoregressive parameter associated with the backshift operator raised to power i, B^i; θ_i is an MA parameter associated with B^i; the back-shift operator B is defined as

$$BX_t = X_{t-1} \tag{23}$$

and α_t is white noise that can be defined as

$$\alpha_t = (X_t - \mu) - \varphi_1(X_{t-1} - \mu). \tag{24}$$

The ARIMA model is the general model, and the autoregressive model and MA model can be viewed as special subclasses of the ARIMA model.

The Box and Jenkins' method was also used to develop an ARIMA model. According to this method, the sample autocorrelation and partial sample autocorrelation diagrams should be examined first to determine the order (p, d, q) of the ARIMA model [25]. The maximum-likelihood method or the conditional least-squares method can then be used to fit the model. However, these methods cannot deal with intermittent GPS data in which second-by-second speed observations are largely missing. Hence, in this study, the ARIMA model was used to forecast the acceleration after the transformation of the irregular speed data measured by time into regularly spaced acceleration data defined by the index of time intervals. The transformation uses two adjacent speed data points to determine the average acceleration of the time interval defined by them, and the average acceleration is set to be equal to the change in speed divided by the length of the time interval. Although the length of each time interval can be different, the index of each interval is discrete in nature and is smaller than the index of the next interval by one. Hence, the acceleration is regularly spaced after the transformation is performed.

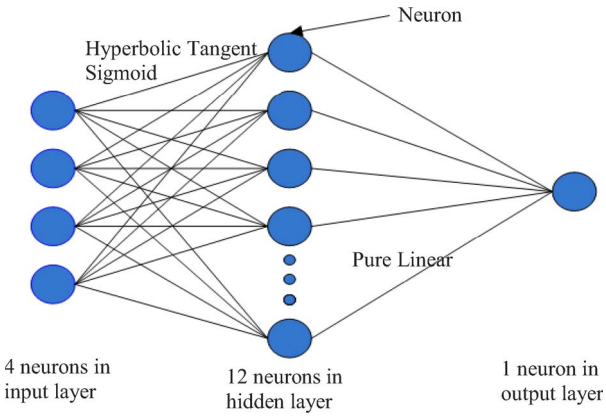

Fig. 4. Structure of the neural network.

B. Utilization of Acceleration Information With a Neural Network

A neural network was used to improve the speed prediction by aggregating the acceleration forecast from the ARIMA models and the speed forecast from the three existing parametric methods. A neural network was used for aggregation for the following three reasons: 1) The major advantage of neural networks is their flexible nonlinear modeling capability [15]. 2) The literature (e.g., [15], [26], and [27]) has shown that the combined method based on neural networks, and ARIMA models can handle both linear and nonlinear time series relationship and can outperform ARIMA models and neural networks alone. 3) Although other nonneural-network-based combined forecasting methods are available for regularly spaced data, they may not be able to handle irregularly spaced data. In this paper, an extrapolation method using acceleration information and a neural network without acceleration information were also included for comparison.

1) Neural Network With Acceleration: A neural network can learn the relationship between inputs and outputs by iteratively adjusting the weights in the network to minimize the mean of the squared difference between the observed inputs and the observed outputs. The neural network used was a feedforward back-propagation network with an input layer of four neurons, a hidden layer of 12 neurons, and an output layer of one neuron. The four neurons in the input layer relate to the four inputs, the forecasting results from the three existing methods, and the acceleration forecasting from the ARIMA model. The single neuron in the output layer represents the estimate for the one-step or multistep ahead forecasting. The number of neurons in the hidden layer was set to 12 because this gave the best performance in a sensitivity test on the number of neurons (not shown here). The hyperbolic tangent sigmoid transfer function was used in the hidden layer, and the pure linear transfer function was used in the output layer because they also give the best performance. The Levenberg-Marquardt back-propagation learning algorithm, which is one of the common and efficient algorithms for training the weights, was chosen to develop the neural network. Fig. 4 shows the structure of the neural network used.

2) Extrapolation Method With Acceleration: The extrapolation method employed the latest observed speed y_t and the predicted acceleration a_t to predict the speed $\hat{y}_{t+k}$, which can be written as

$$\hat{y}_{t+k} = y_t + ka_t. \tag{25}$$

3) Neural Network Without Acceleration: This method aggregated the results of the naive method, the exponential smoothing method, and Holt's method using a neural network. This was then compared with the neural network with acceleration. After a sensitivity test and parameter training were performed, we found that the optimal structure of the neural network without acceleration was similar to that with acceleration, but no acceleration information was included in the network; thus, only three neurons in the input layer were considered. The number of neurons in the hidden layer remained 12.

C. Neural Network Incorporating Information From Adjacent Segments

Up to this point, the study has focused on the correlations within a single segment. However, the external influence of other segments may help improve the forecasting performance, because such influence can be viewed as a result of shockwaves spreading through traffic flow. A neural network was again used to capture and simulate the mechanism of such shockwaves.

Only the directly adjacent segments (one upstream segment and one downstream segment) were selected for analysis, because each segment's length was 100 m, and the effect of a shockwave traveling beyond this distance may not be significant for short-term forecasting (with the length of forecasting time interval being less than 1 min). M3 was used to represent the aggregation of the present and two adjacent segments, and the former neural network was marked as an M1 scheme.

The information from the adjacent segments was carefully chosen because the shockwave from an upstream segment or a downstream segment does not have an immediate effect on the current segment. Shockwaves travel at a certain speed, which suggests that shockwaves observed in an adjacent segment at a certain time will be observed in the current segment at a later time. For example, if a shockwave speed is 1 unit of segment length per unit of time, then a positive acceleration in Segment 3, which is an upstream of Segment 4, at time $t = 0$ may suggest a speed increment in Segment 4 at time $t = 1$, whereas a negative acceleration in Segment 5, which is a downstream of Segment 4, at time $t = 0$ may suggest a jam in Segment 5 and thus a decrease in speed in Segment 4 at time $t = 1$. The neural network was thus revised to adopt this feature.

The new neural network was a feedforward back-propagation network with 24 neurons in the input layer and one neuron in the output layer. The Levenberg-Marquardt back-propagation learning algorithm was again adopted. The 24 neurons represent the three pieces of speed information plus one piece of acceleration information from the three segments (the upstream segment, the current segment, and the downstream segment) at two time indexes (the current time and an earlier time). The single neuron in the output layer represents the estimate for the forecasting.

Four schemes with different earlier times were designed to see which offered the best performance. Depth or time lag is the time difference between the earlier time and the current time and represents the mean time required for a shockwave to travel from an adjacent segment to the current segment. The four schemes were labeled as D1, D5, D10, and D20, where the numbers are seconds.

As the shockwaves from upstream and downstream segments may travel at different speeds, a scheme with different depths for the upstream and downstream segments was also tested. A D20-1-5 scheme was designed to allow for this possibility. It assumed that the downstream information would need 20 s to travel backward and the upstream information would need 5 s to travel forward and that the information from the current segment needed no time to travel; thus, the nearest information (1 s) was used.

For comparison, an M3-D0 scheme with a neural network with 12 neurons in the hidden layer that included information from three segments at the current time only was also developed.

The performance of the schemes was evaluated in a best-to-best manner, whereby the number of neurons that gave the best performance for each scheme was first identified and then adopted for the comparison of the various schemes.

D. MOE

Four measures of effectiveness (MOEs) are commonly used to measure the effectiveness of forecasting methods: 1) the mean absolute relative error (MARE); 2) mean absolute error (MAE); 3) root mean square error (RMSE); and 4) maximum error (ME). These measures compare the forecast results with the observed values in terms of an absolute difference (e.g., MAE and ME), a squared difference (e.g., RMSE), or an absolute relative difference (e.g., MARE). All of the measures help indicate the extent to which the predicted value approaches the true observed value. They offer not only a fair comparison of the performance of different methods but also a standard by which to measure the accuracy of the estimate in relation to the true value. The MARE, MAE, RMSE, and ME are defined as

$$\text{MARE} = \frac{1}{N}\sum_{n=1}^{N}\frac{|X_n - \hat{X}_n|}{X_n} \tag{26}$$

$$\text{MAE} = \frac{1}{N}\sum_{n=1}^{N}|X_n - \hat{X}_n| \tag{27}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(X_n - \hat{X}_n)^2} \tag{28}$$

$$\text{ME} = \max_{n}|X_n - \hat{X}_n| \tag{29}$$

respectively, where X_n is the value observed at time n, and $\hat{X}_n$ is the forecast for time n. In the next section, all of the MOEs were evaluated based on the validation set, i.e., the observations in the final two weeks.

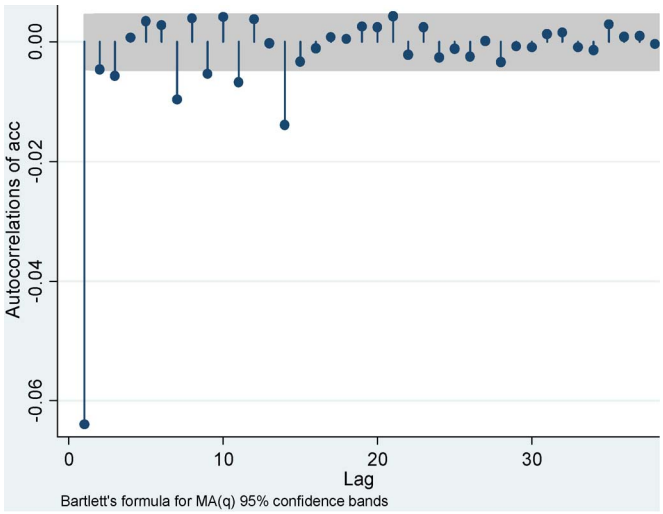

Fig. 5. Autocorrelation diagram of acceleration.

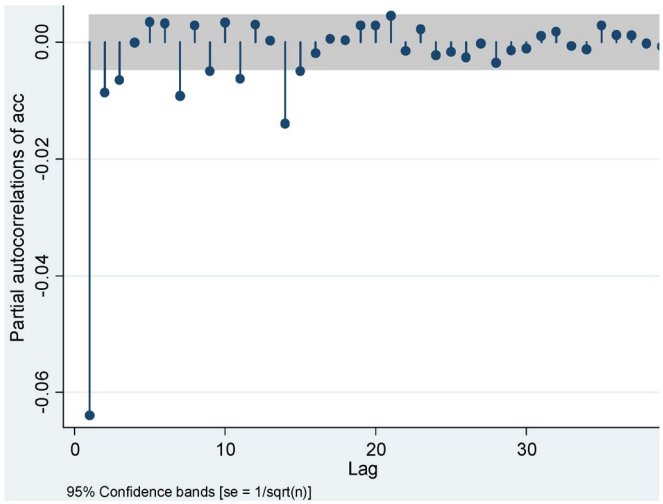

Fig. 6. Partial autocorrelation diagram of acceleration.

IV. Results and Discussion

A. ARIMA Model Order and Parameter Setting

In this section, we focus on 1-s-ahead prediction, unless otherwise specified. The autocorrelation diagram and partial autocorrelation diagram of acceleration were examined to define the order of the ARIMA model.

STATA was used to calculate the autocorrelation diagram and the partial autocorrelation diagram for the sample acceleration. The diagrams are presented in Figs. 5 and 6.

In Fig. 5, for large lags, the sample autocorrelations are all very small and do not follow a very smooth pattern, so there is no need for differencing. The first sample autocorrelation is very large, and the second is slightly bigger than the confidence band, whereas the remainder are quite small, suggesting an MA(2) model for the series.

Similarly, in Fig. 6, the first three sample partial autocorrelations are moderately large, whereas the remainder are very small, suggesting an AR(3) model for the series.

TABLE III
SBC OF THREE ARIMA MODELS WITH DIFFERENT ORDERS

ARIMA (p,0,q)	ARIMA (1,0,1)	ARIMA (3,0,0)	ARIMA (0,0,2)	ARIMA (2,0,1)
CONS	-.00675	-.00675	-.00675	-.00675
AR1	.09435	-.06460		.33287
AR2		-.00909		.01612
AR3		-.00651		
MA1	-.15889		-.06464	-.39750
MA2			-.00526	
SBC	445595.6	445602.8	445596.5	445605.5

Taking these two parameters into consideration, we believe that it may also be useful to include both ARMA (1, 1) and ARMA (2, 1) models in the analysis.

The Schwartz Bayesian criterion (SBC), which is often called the Bayesian information criterion, was used to assess which model best fits the sample series. As shown in Table III, the ARIMA (1, 0, 1) model has the smallest SBC. The ARIMA (1, 0, 1) model was thus selected to make forecasts.

For exponential smoothing and Holt's modification, the parameter(s) is (are) chosen to give to a minimum MARE for the calibration data set.

B. Comparison of Forecast Results Based on a Single Segment

The six forecast methods based on the correlation within a single segment only were used to make a forecast. The MARE and MAE of each segment were calculated for time points at which there were observations. The results are presented in Table IV. This table clearly shows that the MAE value gradually increases when the segments are getting further away from the Cross Harbour Tunnel or when the average traveling speed increases from 5.14 to 14.2 km/h. This seems to indicate that all of the methods perform better when the speed is low.

Due to the congestion effect, most of the segments have low traveling speed. Hence, we use the MARE to determine the best model, as it is more sensitive to speed change at low speed.

Generally speaking, the forecasting methods with acceleration information outperform the methods that include only speed information, indicating the usefulness of acceleration information. Even if we only rely on acceleration information derived by a simple extrapolation relationship, the forecasting results are more accurate in terms of the MARE than those obtained with the complex methods that deal only with speed. The extrapolation method has a MARE of 2.54, whereas the corresponding values for the other three time series forecasting methods are 2.79, 2.76, and 2.71. This indicates the important role of acceleration information in forecasting speed.

The neural network without acceleration produced the worst results, compared with the three parametric methods in terms of the MARE, which shows that a larger number of parameters does not necessarily produce better results. However, in terms of the RMSE, the neural network without acceleration produced the best results, compared with the three parametric methods. This is because we trained the neural network based on minimizing the mean squared error (which is equivalent to minimizing the RMSE) and because the MARE is a relative-error measure that is highly affected by small denominator values, whereas the RMSE is not. In terms of the MAE, the neural network without acceleration still produced the best results, compared with two of the three parametric methods, because the MAE is still not a relative-error measure and its performance is close to, although not exactly the same as, the RMSE.

In fact, a carefully designed scheme that utilizes the important characteristics of the problem is important in the development of a forecasting scheme. This can be verified by the fact that the neural network with acceleration information, which was an aggregation of the three existing parametric methods, outperformed all of the other methods. The comparison of the neural-network methods with and without acceleration information shows that, with almost the same parameters, the presence of the acceleration information greatly increases the forecast precision. Clearly, acceleration information offers a useful indication of the phase transition in the evolution of traffic conditions and thus helps in accurately predicting changes in speed more.

Table IV also shows that the MARE varies across different segments. The MARE for Segment 8 obtained from most of the proposed methods is obviously higher than that for the other segments possibly because this segment involves flyovers and ramps with a large radius, which is substantially different from the geometry of other segments. For Segment 20, the MAE is relatively large because this segment includes a flyover entry.

C. Comparison of Forecast Results Using Data From Adjacent Segments

To account for the spatial correlation between adjacent segments, we extended the neural networks to incorporate results from upstream and downstream segments with different time lags. To give a more comprehensive description of the forecasting performance, we also provided the RMSE and ME in addition to the MAE and MARE for our forecast in Table V, which shows the forecasting results based on the aggregation of the information from adjacent segments. The table shows that all of the schemes, except for the M3-D1 scheme, outperform the M3-D0 scheme in terms of MAE, MARE, and RMSE. This indicates that the consideration of shockwaves from upstream and downstream may help improve the forecasting results. M3-D1 performed worse because 1 s is not enough for a shockwave to travel 100 m. The M3-D10 scheme gives the best performance of all MOEs, which suggests that 10 s is a more appropriate time length than 5 or 20 s for a shockwave to travel along this road section. The asymmetrical scheme in which forward and backward shockwaves were treated differently does not improve the forecasting results. The ME is quite large in all of the schemes. However, these large error values are always obtained when the interval between two consecutive observations is long, e.g., 9000 s (because there are many missing observations).

Table VI provides the results of hypothesis tests on whether the differences in the MAE and MARE between different

TABLE IV
PERFORMANCE OF THE SIX DIFFERENT FORECASTING METHODS (MARE, MAE, AND RMSE)

Segment No.	Existing Parametric Methods			Proposed Methods		
	Naïve	Exponential Smoothing (alpha=0.004)	Holt's Method (Alpha=0.0009 Beta=0.0000004)	Neural Network without Acceleration	Extrapolation Method with Acceleration	Neural Network with Acceleration
1	3.60 (6.71)	3.24 (5.86)	2.50 (5.69)	3.31 (6.60)	3.34 (6.80)	1.78 (5.14)
2	3.42 (8.73)	3.39 (7.82)	3.10 (8.07)	3.70 (7.37)	3.19 (9.08)	2.07 (5.85)
3	2.07 (7.57)	2.09 (6.78)	2.24 (7.14)	2.41 (6.81)	1.74 (8.51)	1.47 (5.43)
4	2.09 (8.15)	2.09 (7.46)	2.17 (7.74)	2.49 (7.13)	1.76 (9.52)	2.35 (5.51)
5	1.38 (8.08)	1.36 (7.39)	1.48 (7.36)	1.61 (6.96)	1.18 (9.00)	0.85 (5.37)
6	2.47 (8.39)	2.35 (7.98)	2.28 (7.96)	2.76 (8.38)	2.42 (9.56)	1.74 (5.99)
7	2.82 (8.15)	2.70 (7.95)	2.55 (7.91)	3.04 (8.42)	2.36 (10.17)	2.13 (6.55)
8	5.80 (9.30)	6.15 (9.10)	6.48 (9.29)	7.30 (9.79)	6.68 (13.14)	9.62 (7.88)
9	5.86 (13.13)	6.13 (13.16)	5.96 (13.93)	7.28 (13.11)	4.89 (17.18)	4.29 (9.81)
10	3.49 (12.16)	4.22 (11.63)	5.77 (12.01)	5.20 (11.53)	3.07 (14.96)	2.79 (7.93)
11	3.24 (10.91)	3.15 (10.39)	3.30 (10.81)	4.19 (10.77)	2.40 (13.41)	2.66 (7.93)
12	2.48 (10.81)	2.55 (10.49)	2.96 (10.97)	3.51 (10.69)	2.06 (13.63)	2.15 (7.67)
13	4.73 (13.28)	4.92 (13.05)	5.61 (13.31)	7.76 (12.27)	3.86 (16.95)	4.39 (9.07)
14	1.14 (10.51)	1.09 (10.15)	1.11 (10.59)	1.28 (10.87)	1.18 (13.99)	0.81 (7.46)
15	1.26 (11.52)	1.52 (11.38)	1.70 (13.07)	1.54 (11.77)	1.28 (16.02)	1.21 (7.94)
16	1.12 (12.12)	1.13 (11.81)	1.37 (12.90)	1.59 (11.24)	1.09 (15.73)	1.09 (8.56)
17	1.20 (12.71)	1.06 (12.10)	0.95 (13.20)	1.24 (11.06)	1.08 (16.82)	0.88 (7.77)
18	0.46 (11.69)	0.44 (11.20)	0.49 (11.19)	0.73 (9.65)	0.51 (16.77)	0.47 (7.42)
19	1.69 (14.46)	1.66 (14.08)	1.86 (13.61)	2.12 (13.06)	1.47 (20.14)	1.15 (8.51)
20	5.01 (23.73)	5.13 (23.17)	5.14 (23.12)	5.64 (20.94)	4.42 (26.76)	2.82 (14.20)
Combined All	2.79 (9.43) *15.03*	2.76 (8.81) *14.26*	2.71 (9.06) *14.46*	3.33 (8.84) *13.39*	2.54 (11.18) *17.21*	1.74 (6.60) *10.41*

* MARE in regular face; MAE in km/h is inside brackets; RMSE in km/h is in italic

TABLE V
PERFORMANCE OF THE DIFFERENT SCHEMES

MOE	M3-D0	M3-D1	M3-D5	M3-D10	M3-D20	M3-D20-1-5
MARE	1.62	1.67	1.52	1.34	1.39	1.46
MAE (km/h)	5.93	5.98	5.47	5.20	5.33	5.54
RMSE (km/h)	9.63	9.10	8.60	8.36	8.50	8.75
ME (km/h)	73.94	72.94	69.31	67.99	71.24	72.83

TABLE VI
MEAN DIFFERENCES BETWEEN DIFFERENT SCHEMES

Schemes	MARE		MAE	
	t-value	p-value	t-value	p-value
M3-D10 Vs M3-D0	2.82	0.0049	14.93	2.59E-50
M3-D10 Vs M3-D1	3.38	0.0007	13.15	2.09E-39
M3-D10 Vs M3-D5	1.93	0.0537	4.62	3.77E-06
M3-D10 Vs M3-D20	0.48	0.6337	2.27	0.0231
M3-D10 Vs M3-D20-1-5	1.26	0.2095	5.81	6.27E-09

schemes are equal to zero or not. This table clearly illustrates that the MAEs of all schemes are statistically different at the significance level of 5%, but only the M3-D0, M3-D1 and M3-D10 schemes have quite different performances statistically in terms of the MARE at the same significance level.

D. Comparison of Forecast Results Based on Information From a Single Segment and Adjacent Segments

The best results from forecasting with the M1 and M3 schemes are compared in Table VII. The results suggest that a substantial improvement was achieved by aggregating the information from adjacent segments in terms of all the MOEs. The acceleration information from adjacent segments may account for this improvement, as this is a direct description of phase-transition effects on speed forecasting. The acceleration information from adjacent segments may be more efficient in identifying phase-transition effects than the acceleration information in the current segment.

E. Discussion of the Forecasting Results

The best MARE in the study was 1.34, which seems large. However, it can be seen from (26) that the MARE tends to overweigh the relative error from the small observed values.

TABLE VII
COMPARISON OF M1 AND M3-D10

MOE	M1 (Neural Network with Acceleration)	M3-D10
Best MARE	1.74	1.34
Best MAE (km/h)	6.60	5.20
Best RMSE (km/h)	10.41	8.36
Best ME (km/h)	74.66	67.99

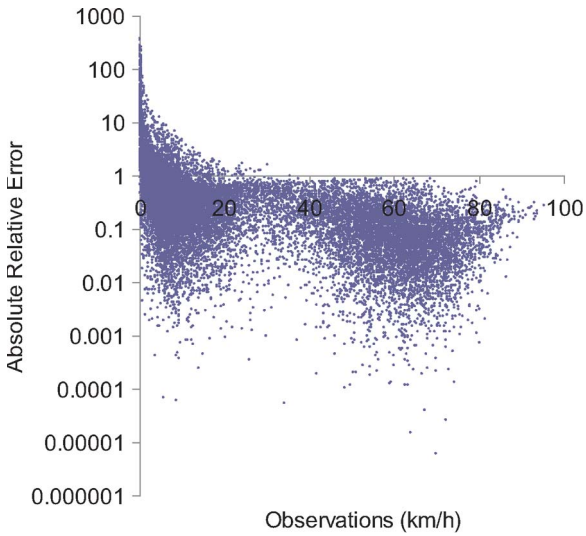

Fig. 7. Logarithmic absolute relative error versus observations.

A small absolute forecasting error would result in a larger relative error for low speeds than for high speeds. As traffic jams frequently occur on the studied road section, the speeds are often low. Hence, the large absolute relative error from the low speed dominates the MARE.

Fig. 7 shows the distribution of the absolute relative errors of the best performing M3-D10 scheme. A logarithmic scale was used for the vertical axis. From this figure, we can see that, when the observed speed is very small, the absolute relative error is often very large. However, when the observed speed is more than 20 km/h, most of the relative errors are less than 1. As most of the speeds are smaller than 20 km/h, the MARE is large.

We report the MAE and RMSE in Table VII to demonstrate the performance of our forecasting methods. The MAE and RMSE measure the average absolute errors from forecasts in relation to observations. A comparison of the MAE and RMSE values generates the same conclusion as that derived from the MARE. The best performer, i.e., the M3-D10 scheme, had a MAE of only 5.20 km/h and an RMSE of 8.36 km/h, despite a seemingly large MARE of 1.34. The MAE of 5.20 km/h means that the predicted speed is about ± 5 km/h on average, which is sufficiently accurate for applications such as identifying the occurrence of congestion and phase transitions, as well as incident detection.

Fig. 8 shows a comparison of the forecast and observed speeds based on the M3-D10 scheme for the validation set,

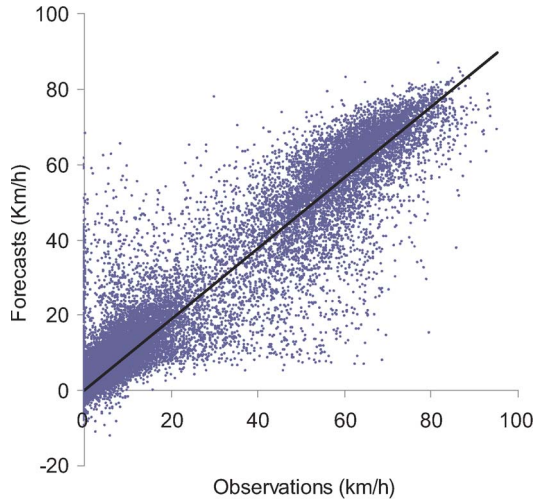

Fig. 8. Forecasts versus observations.

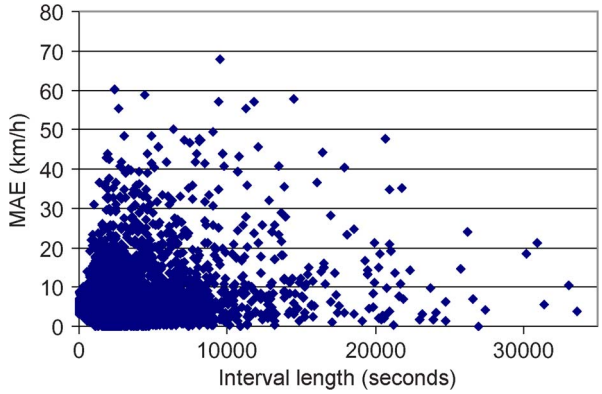

Fig. 9. MAE sorted by interval length.

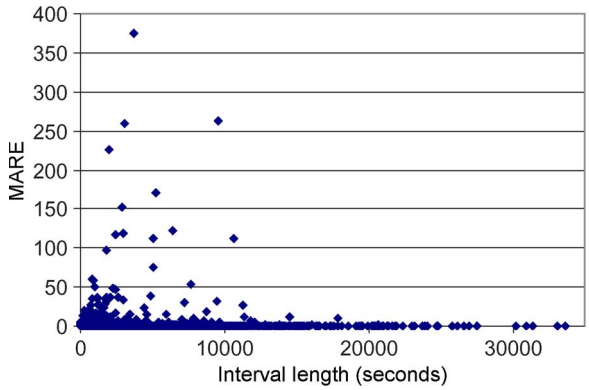

Fig. 10. MARE sorted by interval length.

which further illustrates that the proposed method offers a good fit with the observed data. The correlation coefficient of the forecasts and observations in Fig. 8 is 0.9435, which implies that there is good consistency between the forecasts and the observations.

To obtain a more comprehensive understanding of the effect of the degree of missing observations on the accuracy of our forecasting method, we present the results of M3-D10 in Fig. 9 with the MAE sorted by the length of time intervals, in which the length reflects the number of observations missing minus

TABLE VIII
MULTISTEP-AHEAD FORECASTING RESULTS

k	Naïve		M1		M3-D10	
	MARE	MAE (km/h)	MARE	MAE (km/h)	MARE	MAE (km/h)
1 second	2.79	9.43	1.74	6.6	1.34	5.2
5 seconds	2.81	9.45	2.11	12.47	1.47	13.88
10 seconds	2.82	9.47	2.11	13.33	1.50	15.08
15 seconds	2.82	9.48	2.11	13.67	1.50	15.52
30 seconds	2.83	9.51	2.11	14.18	1.59	16.10
60 seconds	3.29	9.99	2.03	15.62	1.59	17.52

one. In general, when the interval length is short (e.g., less than 800 s), the accuracy decreases with the increase in the interval length because using more recent observations results in better prediction. However, when the length is long (i.e., longer than 800 s, the missing observation does not have a significant impact on the prediction accuracy, as the traffic condition used to predict the traffic condition in the future has no relationship with each other. Fig. 10 shows the extent of the effect of missing observations on the MARE, which agrees with the results presented in Fig. 9.

Table VIII compares the M1 scheme, the M3-D10 scheme and the naive method for various values of k. This table illustrates that both the MAE and the MARE increase as the value of k increases. The naive method always gives the largest MARE. M3-D10 performs better than M1 in terms of MARE but worse in terms of MAE. The largest MAE in this table is still acceptable for applications such as identifying the occurrence of congestion and phase transitions, as well as incident detection.

V. CONCLUSION

The study reported here has employed several methods to forecast traffic speeds using irregularly spaced GPS data from taxis located on a two-kilometer section of Princess Margaret Road approaching the Cross Harbour Tunnel in Hong Kong. Three existing methods, i.e., the naive method, the simple exponential method, and Holt's method, have been used to make forecasts using the irregularly spaced GPS data. Neural network models have been used to aggregate the speed forecasting results from the three methods, and the acceleration forecasting result from an ARIMA model has been used to improve the forecasting performance. Information from adjacent segments has also been included in the neural networks to further improve the forecasting accuracy. Six schemes, i.e., M3-D0, M3-D1, M3-D5, M3-D10, M3-D20, and M3-D20-1-5, which utilized the information from adjacent segments in different ways, have been tested. The MARE has been chosen to measure and compare the forecasting performance of the methods. The results have shown that the neural network with acceleration information and information from adjacent segments outperformed all of the other methods. The inclusion of acceleration information improves the neural network forecasting results because the resulting forecasting utilizes information from irregular time intervals and captures phase transitions for speed forecasting. The use of information from adjacent segments further improves the forecasting results, because it provides a better description of the phase-transition effects.

In the future, the proposed forecasting method will be combined with the existing forecasting methods, which rely on fixed-location detection infrastructure speed data, to improve the forecasting accuracy at those fixed locations. Moreover, we will also consider a model-based or fuzzy-neural approach to enhance the forecasting performance by giving a theoretical characterization of phase transitions in the forecasting procedure (e.g., see [14] and [28]).

ACKNOWLEDGMENT

The authors would like to thank Concord Pacific Satellite Technologies Ltd. and Motion Power Media Ltd. for kindly providing GPS data from the taxis. The authors are grateful to the constructive comments of the three reviewers.

REFERENCES

[1] J. D. Croston, "Forecasting and stock control for intermittent demands," *Oper. Res. Quart.*, vol. 23, no. 3, pp. 289–303, Sep. 1972.
[2] R. H. Jones, "Maximum-likelihood fitting of ARMA models to time-series with missing observations," *Technometrics*, vol. 22, no. 3, pp. 389–395, Aug. 1980.
[3] D. J. Wright, "Forecasting data published at irregular time intervals using an extension of Holt method," *Manage. Sci.*, vol. 32, no. 4, pp. 499–510, Apr. 1986.
[4] T. Cipra, J. Trujillo, and A. Rubio, "Holt-Winters method with missing observations," *Manage. Sci.*, vol. 41, no. 1, pp. 174–178, Jan. 1995.
[5] J. L. Carmo and A. J. Rodrigues, "Adaptive forecasting of irregular demand processes," *Eng. Appl. Artif. Intell.*, vol. 17, no. 2, pp. 137–143, Mar. 2004.
[6] T. R. Willemain, C. N. Smart, and H. F. Schwarz, "A new approach to forecasting intermittent demand for service parts inventories," *Int. J. Forecast.*, vol. 20, no. 3, pp. 375–387, Jul.–Sep. 2004.
[7] W. F. Velicer and S. M. Colby, "A comparison of missing-data procedures for ARIMA time-series analysis," *Educ. Psychol. Meas.*, vol. 65, no. 4, pp. 596–615, Aug. 2005.
[8] T. Cipra, "Exponential smoothing for irregular data," *Appl. Math.*, vol. 51, no. 6, pp. 597–604, Dec. 2006.
[9] N. Altay, F. Rudisill, and L. A. Litteral, "Adapting Wright's modification of Holt's method to forecasting intermittent demand," *Int. J. Prod. Econ.*, vol. 111, no. 2, pp. 389–408, Feb. 2008.
[10] T. Cipra and T. Hanzak, "Exponential smoothing for irregular time series," *Kybernetika*, vol. 44, no. 3, pp. 385–399, 2008.
[11] B. L. Smith and M. J. Demetsky, "Traffic flow forecasting: Comparison of modeling approaches," *J. Transp. Eng. ASCE*, vol. 123, no. 4, pp. 261–266, Jul./Aug. 1997.
[12] B. L. Smith, B. M. Williams, and R. Keith Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transp. Res. C, Emerging Technol.*, vol. 10, no. 4, pp. 303–321, Aug. 2002.
[13] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng. ASCE*, vol. 129, no. 6, pp. 664–672, Nov./Dec. 2003.
[14] W. Y. Szeto, B. Ghosh, B. Basu, and M. OMahony, "Multivariate traffic forecasting technique using cell transmission model and SARIMA model," *J. Transp. Eng. ASCE*, vol. 135, no. 9, pp. 658–667, Sep. 2009.

[15] M. C. Tan, S. C. Wong, J.-M. Xu, Z.-R. Guan, and P. Zhang, "An aggregation approach to short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 1, pp. 60–69, Mar. 2009.
[16] A. Simroth and H. Zähle, "Travel time prediction using floating car data applied to logistics planning," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 243–253, Mar. 2011.
[17] S. Sun and X. Xu, "Variational inference for infinite mixtures of Gaussian processes with applications to traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 466–475, Jun. 2011.
[18] A. Khosravi, E. Mazloumi, S. Nahavandi, D. Creighton, and J. W. C. van Lint, "Prediction intervals to account for uncertainties in travel time prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 537–547, Jun. 2011.
[19] Y. Wang, P. Coppola, A. Tzimitsi, A. Messmer, M. Papageorgiou, and A. Nuzzolo, "Real-time freeway network traffic surveillance: Large-scale field-testing results in Southern Italy," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 548–562, Jun. 2011.
[20] F. Soriguera and F. Robusté, "Highway travel time accurate measurement and short-term prediction using multiple data sources," *Transportmetrica*, vol. 7, no. 1, pp. 85–109, Jan. 2011.
[21] K. Y. Chan, T. S. Dillon, J. Singh, and E. Chang, "Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and Levenberg–Marquardt algorithm," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 644–654, Jun. 2012, DOI:10.1109/TITS.2011.2174051.
[22] T. T. Tchrakian, B. Basu, and M. O'Mahony, "Real-time traffic flow forecasting using spectral analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 519–526, Jun. 2012, DOI:10.1109/TITS.2011.2174634.
[23] H. Chen, S. Grant-Muller, L. Mussone, and F. Montgomery, "A study of hybrid neural network approaches and the effects of missing data on traffic forecasting," *Neural Comput. Appl.*, vol. 10, no. 3, pp. 277–286, Dec. 2001.
[24] J. W. C. van Lint, S. P. Hoogendoorn, and H. J. Van Zuylen, "Accurate freeway travel time prediction with state-space neural networks under missing data," *Transp. Res. C*, vol. 13, no. 5/6, pp. 347–369, Oct. 2005.
[25] G. E. P. Box, *Time Series Analysis; Forecasting and Control*, G. E. P. Box and G. M. Jenkins, Eds. San Francisco, CA: Holden-Day, 1970.
[26] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, Jan. 2003.
[27] Y. Gao and M. J. Er, "NARMAX time series model prediction: Feedforward and recurrent fuzzy neural network approaches," *Fuzzy Sets Syst.*, vol. 150, no. 2, pp. 331–350, Mar. 2005.
[28] H. B. Yin, S. C. Wong, J. Xu, and C. K. Wong, "Urban traffic flow prediction using a fuzzy-neural approach," *Transp. Res. C, Emerging Technol.*, vol. 10, no. 2, pp. 85–98, Apr. 2002.

Qing Ye received the B.Eng. degree from Tsinghua University, Beijing, China, and the M.Phil. degree from the University of Hong Kong, Pokfulam, Hong Kong.

He is currently with the Department of Civil Engineering, University of Hong Kong. His research interests include traffic-flow-forecasting methods.

W. Y. Szeto received the B.Eng., M.Phil., and Ph.D. degrees from Hong Kong University of Science and Technology, Kowloon, Hong Kong.

He is currently an Assistant Professor with the Department of Civil Engineering, University of Hong Kong, Pokfulam, Hong Kong. His research interests include dynamic traffic assignment, transport network design, transport network reliability, transport logistics, public transport, metaheuristics, game theory, traffic flow forecasting, and sustainable transport.

S. C. Wong received the B.Sc.(Eng) and M.Phil. degrees from the University of Hong Kong, Pokfulam, Hong Kong, and the Ph.D. degree from University College London, London, U.K.

He is currently a Chair Professor with the Department of Civil Engineering, University of Hong Kong. His research interests include optimization of traffic signal settings, continuum modeling for traffic equilibrium problems, land use and transportation problems, dynamic highway and transit assignment problems, urban taxi services, and road safety.