Short-Term Forecasting of Traffic Volume

Evaluating Models Based on Multiple Data Sets and Data Diagnosis Measures

Lei Lin, Qian Wang, and Adel W. Sadek

Although several methods for short-term forecasting of traffic volume have recently been developed, the literature lacks studies that focus on how to choose the appropriate prediction method on the basis of the statistical characteristics of the data set. This study first diagnosed the predictability of four traffic volume data sets on the basis of various statistical measures, including (*a*) complexity analysis methods, such as the delay time and embedding dimension method and the approximate entropy method; (*b*) nonlinearity analysis methods, such as the time reversibility of surrogate data; and (*c*) long-range dependency analysis techniques, such as the Hurst exponent. After the data sets were diagnosed, three models for short-term prediction of traffic volume were applied: (*a*) seasonal autoregressive integrated moving average (SARIMA), (*b*) *k* nearest neighbor (*k*-NN), and (*c*) support vector regression (SVR). The results from the statistical data diagnosis methods were then correlated to the performance results of the three prediction methods on the four data sets to determine the means for choosing the appropriate prediction method. The results revealed that SVR was more suitable for nonlinear data sets, while SARIMA and *k*-NN were more appropriate for linear data sets. The data diagnosis results were also used to devise a selection process for the parameters of the prediction models, such as the length of the training data set for SARIMA and SVR, the average number of nearest neighbors for *k*-NN, and the input vector length for *k*-NN and SVR.

The ability to provide short-term forecasts of traffic flow parameters has long been regarded as a key component of advanced traffic management and control system applications. In the past few decades, a variety of prediction models have been developed for that purpose. Generally speaking, these methods can be categorized into parametric methods and nonparametric approaches. Among the most popular parametric models are time series analysis methods, such as the seasonal autoregressive integrated moving average model (SARIMA). For example, Williams and Hoel presented the theoretical basis for modeling univariate traffic condition data streams as SARIMA processes (*1*). Smith et al. showed that SARIMA performs better than the nonparametric nearest-neighbor method for the single-point traffic prediction problem (*2*). Cools et al. used both ARIMA with explanatory variables and SARIMA with explanatory variables to predict daily traffic volumes (*3*). A major limitation of SARIMA time series analysis in general, however, is that the models assume linear correlation structures among time series data and, thus, the models may not be able to capture the nonlinearity inherent in real-time traffic data.

Nonparametric methods, by contrast, attempt to identify historical data that are similar to the prediction instant and use the average of the identified data items to forecast the future. Nonparametric methods do not rely on predetermined relationship functions between the past and the present and are thus supposedly able to deal with the nonlinearity and nonstationarity of traffic time series. The typical nonparametric methods include computational intelligence (CI) techniques; for example, many different types of neural networks have been proposed (*4–5*). Besides neural networks, support vector regression (SVR) has also been used (*6, 7*). In addition to CI, another popular class of nonparameter models is nearest neighbor methods (*8*).

Although there is an extensive literature on short-term traffic volume prediction, most of the previous studies considered only one modeling technique and a single data set. Even among the comparative studies in the literature, the focus has typically been on comparing the performance of multiple models on a single data set (*9, 10*). The risk of using one data set to test models is that the conclusions derived may be specific to the data set considered. This has often led to inconsistent conclusions among the different studies regarding which modeling method is superior. In addition, single-data–based testing cannot address the essential question that is of particular interest to practitioners (i.e., how to select prediction models based on the data).

Recently, a handful of researchers have begun to pay more attention to that issue. For example, Smith and Demetsky tested four prediction methods on two data sets (*8*). However, the two data sets came from sites on the same highway and there was no discussion in the study about the relationship between the attributes of a data set and model performance. Other researchers have pointed out the importance of data diagnosis before model selection and proposed different measures to indicate data characteristics. For instance, Vlahogianni et al. (*11*) discussed some statistical methods for detecting nonlinearity and nonstationarity of traffic volume time series and Shang et al. (*12*) discussed the nonlinearity property of traffic volumes based on chaos theory. However, no effort was made in those studies to link data diagnosis results with model selection.

In this context, the study reported here used multiple data sets to conduct comprehensive testing of the performance of online traffic prediction models. Three popular prediction models, SARIMA, SVR, and *k* nearest neighbor (*k*-NN), were chosen as representatives of parametric and nonparametric classes. In addition, four data

Department of Civil, Structural, and Environmental Engineering, University at Buffalo, State University of New York, 231 Ketter Hall, Buffalo, NY 14260. Corresponding author: A. W. Sadek, asadek@buffalo.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2392, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 40–47.
DOI: 10.3141/2392-05

sets were used as the test cases. The following data sets were used: (*a*) an hourly traffic volume data set for an international border crossing; (*b*) an hourly traffic volume data set for Interstate 90 (I-90) in Buffalo, New York; (*c*) a 5-min volume data set from Virginia; and (*d*) a 2-min volume data set from Beijing, China. At the core of the testing was the goal of seeking appropriate prediagnosis procedures and statistical measures of volume time series to facilitate the selection of prediction models. A second objective of the study was to utilize the insight gained from the data diagnosis procedures in selecting appropriate parameters for the prediction models based on the properties of traffic volume data.

METHODOLOGY

This section will first briefly introduce the statistical measures used to diagnose traffic volume time series data and then discuss the three prediction methods used, SARIMA, *k*-NN, and SVR.

Statistical Measures for Data Diagnosis

The goal of data diagnosis is to assess the predictability of a time series and to identify which methods are more appropriate for prediction. Multiple measures were used in this study, including: (*a*) measures of complexity, such as delay time and embedding dimension analysis and approximate entropy (ApEn); (*b*) nonlinearity indicators, such as the time reversibility of surrogate data; and (*c*) measures of long-range dependence, such as the Hurst exponent. Given space limitations, only highlights of each measure will be briefly introduced.

Complexity Measures

Delay Time and Embedding Dimension The idea behind the delay time and embedding dimension method is to make a time-delay reconstruction of the phase or state space of the time series in which to view the dynamics of the system (*13, 14*). For a time series $\{x_t\}$, a new time series, denoted by $\{y_t\}$, $y_t = \{x_t, x_{t-\tau}, \ldots, x_{t-(m-1)\tau}\}$, is first constructed. The new space, which consists of such vectors y_t, is called the phase space or state space. The elements in y_t include $(m - 1)$ relevant past values of x_t and the relevant past values may lag τ time intervals from each other. Here, τ and m are called the delay time (or lag value) and embedding dimension, respectively. Typically, the best value of delay is determined by the mutual information method, which seeks to maximize the joint probability $p(X(t), X(t + \tau))$ given τ (*15*). For determining the embedding dimension, the false nearest-neighbor algorithm can be used. The algorithm scans potential values of m in order to identify the optimal value that avoids the inclusion of false or irrelevant data the most (*16*).

Approximate Entropy ApEn is a technique for measuring the magnitude of irregularity or unpredictability of fluctuations in a time series. Specifically, the measure denotes the likelihood that fluctuation patterns of a series do not repeat over time. Small values of ApEn usually indicate a predictable data set with repetitive patterns, whereas larger values of ApEn indicate more randomness. After phase space reconstruction, the analyst needs to specify the threshold for the similarity criterion, r, which defines whether two patterns are similar. Ho et al. provide details on how ApEn is calculated (*17*).

Nonlinear Indicators: Time Reversibility of Surrogate Data

The method of surrogate data tests the nonlinearity of time series by verifying whether a series is consistent with the null hypothesis of a linear Gaussian process. A process known as the iterated amplitude adjusted fourier transform is first used to generate surrogate data sets from the original time series (*18*). This process is repeated several times and, after each time, a measure known as the time reversibility, r, which measures the asymmetry of a series under time reversal, is calculated (*18, 19*). The time reversibility value for the original time series, r_0, is also calculated. Finally, the test checks to see whether r_0 is within the distribution of r. If it is, the original time series is linear; otherwise, it is nonlinear (*20*).

Long-Range Dependence Indicators: Hurst Exponent

The Hurst exponent is a measure to characterize the long-range dependence (LRD) of a time series (*21*). In the time domain, LRD manifests as a high degree of correlation between distantly separated data points. The values of the Hurst exponent range from 0 to 1 and can be categorized into three groups with different implications: (*a*) Hurst = 0.5 implies a random time series; (*b*) 0 < Hurst < 0.5 indicates a trend-reverting tendency by which the increasing (or decreasing) trend observed at present is likely to flip at the next time instant; and (*c*) 0.5 < Hurst < 1 indicates a trend-reinforcing tendency by which the increasing (or decreasing) trend at present is likely to be maintained in the near future. Some research has shown that for back propagation neural network models, time series with a large Hurst exponent can be predicted more accurately than series with a Hurst exponent close to 0.5 (*22*). Hurst provides greater details on how to calculate the Hurst exponent (*23*).

Short-Term Traffic Volume Prediction Models

SARIMA Model

The SARIMA model is one of the most popular and extensively used time series models that exhibit seasonal trends. By following the three steps of identification, estimation, and diagnosis, SARIMA models can be fitted to stationary or weakly stationary time series data. Box et al. provide a detailed discussion of the SARIMA model (*24*).

k-*NN Model*

The *k*-NN model is a prediction method that decides the output by finding the *k*-NN (i.e., most similar) of the input in a historical data set and uses their observed output (i.e., the predicted volume). The Euclidean distance is typically used to assess similarity. When *k*-NNs are found and assuming their corresponding output values are v_i, $i = 1, 2, \ldots, k$, the predicted value (v) can be determined by calculating the weighted average of the neighbors as follows:

$$v = \frac{1}{k}\sum_{i=1}^{k} v_i \qquad (1)$$

Support Vector Regression Model

The support vector machine (SVM) is a popular machine-learning method based on statistical learning theory developed by Vapnik in 1995 (*25*). SVMs were developed to solve classification problems. Later, SVM was extended to allow for solving nonlinear regression problems and this extension resulted in what is known as SVR (*26*). SVR captures nonlinearity of the time series by mapping the training samples into a high-dimensional feature space induced by a kernel function, followed by linear regression in that space. A radial basis function,

$$K(x_i, x_j) = \exp\left(-\gamma \left|x_i - x_j\right|^2\right)$$

is one common kernel function. The parameters of the SVR models, namely the penalty factor C and the gamma in the kernel function, are often optimized with the k-fold cross-validation method (*27*).

MODELING DATA SETS AND DATA DIAGNOSIS RESULTS

The volume data sets chosen for testing purposes in this study represent facilities with different characteristics (e.g., an international border versus a commuter freeway), different locations (New York, Virginia, and Beijing), as well different time resolutions (hourly versus 5 min versus 2 min). Specifically, Data Set 1 came from I-90 in Buffalo, New York (detector M4183E) and is an hourly volume data set. Data Set 2 came from the Peace Bridge international border crossing connecting western New York and southern Ontario; Data Set 2 is also an hourly volume data set. Data Set 3 came from the westbound direction of I-64 in the Hampton Roads area in Virginia; it has a resolution of 5 min. Finally, Data Set 4 is from the second ring road in Beijing, China (Detector 02024) and has a 2-min resolution. The lengths of the time series sets used in the study were 2,000 observations for the hourly volume sets, 5,760 for the 5-min data set, and 3,600 for the 2-min Beijing data set.

Before the model development was applied, the predictability of the four data sets was diagnosed with the statistical measures previously discussed. The details are shown below.

TABLE 1 Mutual Information Values, Time Delay

Data Set	Mutual Information Value by τ Value				
	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 5$
I-90	9.08	9.07	9.08	9.07	9.07
Peace Bridge	6.59	6.58	6.57	6.57	6.58
Virginia	2.77	2.60	2.50	2.42	2.36
Beijing	4.37	4.23	4.11	4.01	3.93

Delay Time and Embedding Dimension Method

The mutual information method was first used to determine the best value for the time delay, τ, for the four data sets. The results are shown in Table 1, which lists the mutual information value for values of τ ranging from 1 to 5. For all four data sets, the best value of τ appears to be 1 because it is the value that corresponds to the maximum mutual information value (*28*).

To determine the value for the embedding dimension, m, Figure 1 plots the percentage of the false nearest neighbors as a function of the embedding dimension values for the four data sets. The best

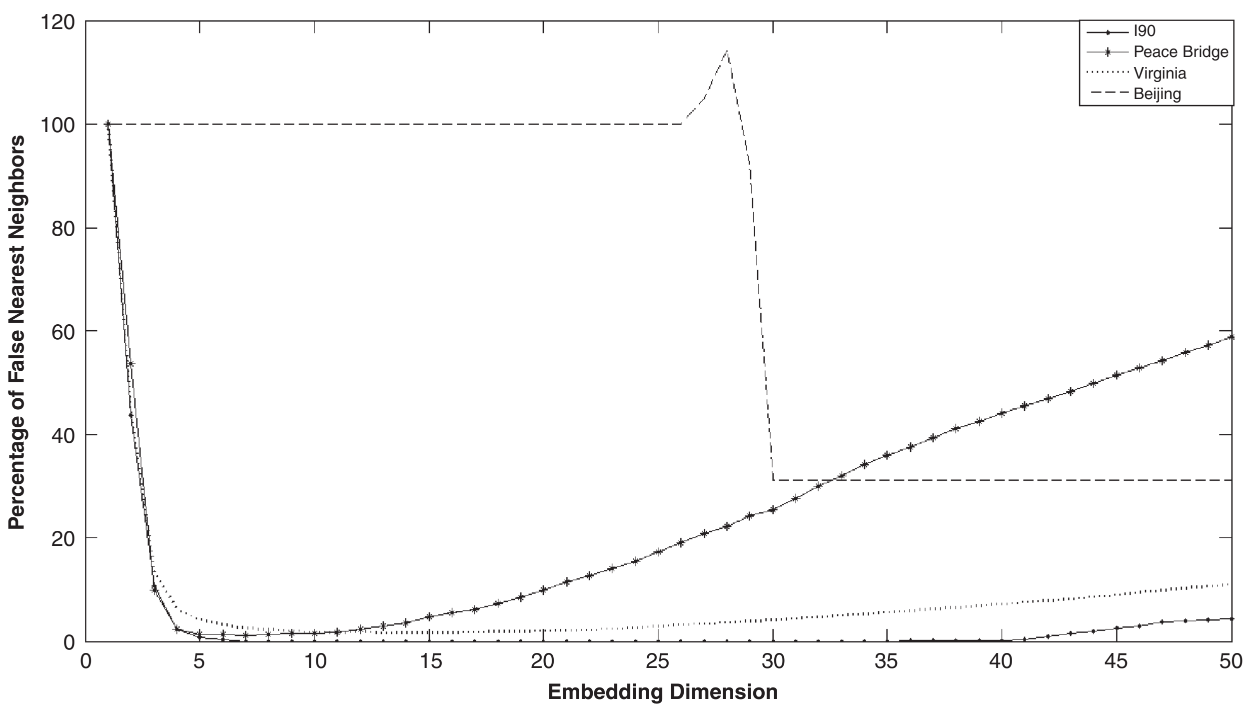

FIGURE 1 Percentage of false nearest neighbors as function of embedding dimension.

values for the embedding dimension appear to be 7, 6, 9, and 30 for the I-90, Peace Bridge, Virginia, and Beijing data sets, respectively, because those values lead to the lowest percentage of false nearest neighbors. Determining the embedding dimension, m, serves the very important role of determining the length of the input vector for the k-NN and SVR prediction methods.

ApEn Method

For each data set, the time delay (τ) and embedding dimension (m) values that were calculated as described in the previous section were used; that is, $\tau = 1$ for all the data sets and $m = 7$ for the I-90 data, $m = 6$ for the Peace Bridge data, $m = 9$ for the Virginia data, and $m = 30$ for the Beijing data. With these values, a new state space was built and, assuming that if the Euclidean distance between two vectors is lower than 20% of the standard deviation of the data set the two patterns are similar, the ApEn for each data set was calculated. The results were an ApEn value of 0.24 for I-90, 0.30 for Peace Bridge, 0.21 for Virginia, and 0.0049 for Beijing.

Given that the approximate entropy is a measure of the predictability or irregularity of the data set (with larger values indicating a time series that is more difficult to predict), the Peace Bridge data were the most unpredictable with the least chance of having repetitive patterns, followed by I-90, Virginia, and finally Beijing, which was the easiest to predict. The outcomes were in agreement with what was expected, because the Peace Bridge data set and the I-90 data set were hourly volumes, whereas the Virginia data set was a 5-min count and the Beijing data set was a 2-min count (naturally it is much more difficult to predict longer times in the future). Moreover, traffic at a border crossing was expected to be more irregular compared with a commuter freeway.

Time Reversibility of Surrogate Data

For each of the four data sets, 100 replicates of surrogate data were generated and for each surrogate data set, the time reversibility value, r, was calculated and a distribution of r was generated, as can be seen in Figure 2. The time reversibility value of the original time series, r_0, was also calculated (this is the location of the straight vertical line in each of the plots in Figure 2).

The I-90 data set and the Virginia data set appear to exhibit linear patterns (since r_0 lies within the distribution of r), whereas the Peace Bridge data set and the Beijing data set exhibit nonlinearity.

Hurst Exponent

The calculated values for the Hurst exponent for the four data sets were as follows: $\text{Hurst}_{\text{I-90}} = 0.26$, $\text{Hurst}_{\text{PB}} = 0.60$, $\text{Hurst}_{\text{Virginia}} = 0.69$, and $\text{Hurst}_{\text{Beijing}} = 0.91$. Because the closer the value of the exponent is to 0.5, the more random the data are, the results seem to indicate that the Peace Bridge data are the most random while the Beijing data are the most stable. Moreover, given that the Hurst value for the I-90 data set is between 0 and 0.5, the data appear to exhibit a trend-reverting tendency. In contrast, the Virginia data and the Beijing data exhibit a trend-reinforcing tendency (Hurst exponent within the range of 0.5 to 1).

MODEL DEVELOPMENT AND CALIBRATION

Following the characterization of each data set, the three prediction methods (SARIMA, k-NN, and SVR) were used to provide short-term forecasts for each of the four test data sets. Following the calibration of

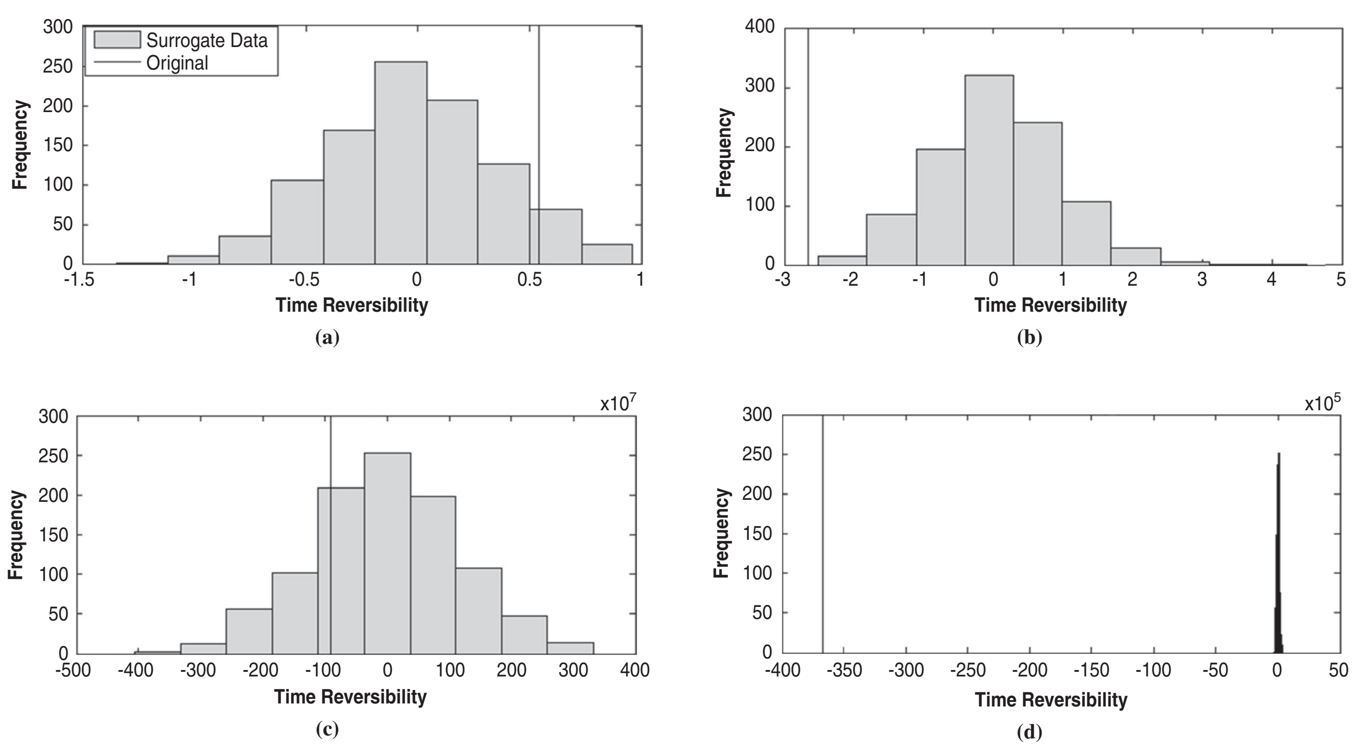

FIGURE 2 Test of nonlinearity through surrogate data: (*a*) I-90, (*b*) Peace Bridge, (*c*) Virginia, and (*d*) Beijing.

each prediction method, its performance on the different data sets was evaluated and correlated to the characteristics of the set as quantified by the data diagnosis measures described above. By doing so, the study hoped to glean useful insight into how to select the best prediction method for a data set given its statistical characteristics. Furthermore, the insights gained from the data diagnosis measures were utilized to guide the design and calibration of the prediction methods.

Development of SARIMA Models

The Statistical Package for Social Sciences was used to build the SARIMA models. An essential step in developing the SARIMA models was to determine the appropriate training data size for each data set. For traffic volume SARIMA models, the appropriate seasonal period is 1 week, and therefore for the hourly data sets (i.e., I-90 and Peace Bridge), a seasonal period of 168 intervals was adopted (i.e., 24 × 7). For the 5-min volume and 2-min volume data sets, considering that the training data set would be very large if the 1-week seasonal period was assumed, the seasonal period was set at 1 day and the models were used to attempt to predict traffic volumes for weekdays only. By excluding the weekends, and by assuming that weekdays were similar to one another, a seasonal period of 1 day would be adequate. With this assumption, the seasonal period was assumed to be equal to 288 intervals (24 × 12) for the Virginia data set and 720 intervals (24 × 30) for the Beijing data set.

In developing the models, various training data set sizes were tested and the value of the resulting mean absolute percent error (MAPE) was monitored (the size tested was always an integer multiple of the assumed seasonal period). To adapt SARIMA to online prediction, the training data set was updated at each step by adding the most recent observation and deleting the oldest. The results are shown in Table 2.

For the I-90 data, the prediction accuracy of SARIMA improved with the increase in the training data size; the best performance was reached with a size equal to 1,008 observations. This was also generally the trend for the other three data sets, although the best performance was achieved at slightly smaller-size training sets (i.e., 840 for Peace Bridge, 864 for Virginia, and 720 points for Beijing). Correlating these observations to the results of the statistical measures performed on the data sets led to some interesting observations. For example, because the Beijing data had a very low approximate entropy value (0.02), it was much easier to predict and required the least number of data points for training. The Hurst exponent value for the Beijing data set was 0.91 (close to 1.0), an indication of a much stronger trend-reinforcing tendency compared with the Hurst values for Virginia and Peace Bridge.

The prediction results also revealed important information about the performance of SARIMA. The SARIMA models performed well for the I-90 data, the Virginia data, and the Beijing data; this performance resulted in acceptable MAPEs lower than 10% in general. This result was not the case for the Peace Bridge data, for which the lowest MAPE was around 18%. The inconsistent performances can be explained by the nonlinearity and less predictability of the Peace Bridge data as identified by the surrogate data and approximate entropy analyses. Theoretically speaking, SARIMA models are built on the linearity assumption and thus may not perform well for nonlinear time series, such as the Peace Bridge data.

Development of *k*-NN Models

Two attributes needed to be specified for the *k*-NN model development: the length of the input vector (*B*), which refers to how many previous time steps were used to predict the next value, and the number of nearest neighbors (*k*). Various combinations of *B* and *k* were tested to see which one would lead to the best performance for each data set. As shown in Figure 3, the best (*B, k*) value combinations that returned the least prediction errors were (7, 3), (24, 3), (11, 1), and (30, 1) for the I-90, Peace Bridge, Virginia, and Beijing data, respectively. How did these values for *B* and *k* correlate with the values of the statistical data diagnosis measures calculated for the data sets?

With respect to *k*, the resulting values seemed to correlate well with the values for the approximate entropy. The I-90 and Peace Bridge data sets had higher approximate entropies than the Virginia and Beijing data sets, which means that the probability that the former group of data sets would have more different patterns than observed patterns was higher than that for the latter group. As a result, more nearest neighbors may be needed to lower the risk of use of a different pattern for prediction. In terms of the input vector length (*B*), the values were identical (or very close) to the embedding dimension, *m*, values for the I-90, Virginia, and Beijing data sets, where the analysis indicated values for *m* equal to 7 for I-90, 9 for Virginia, and 30 for Beijing. However, this was not the case for the Peace Bridge data. For Peace Bridge, the optimal value of 24 is interesting because this value can be easily explained by the strong seasonality of the data (i.e., the periodical variations of hourly volumes within the 24-h cycle).

Development of SVR Models

The development of the SVR models required the specifications of the training data size and input data vector length (*D*). The simplest

TABLE 2 Performance of SARIMA, Training Data Size

Training Data Set Size	MAPE (%)	Training Data Set Size	MAPE (%)	Training Data Set Size	MAPE (%)	Training Data Set Size	MAPE (%)
I-90		Peace Bridge		Virginia		Beijing	
168	17.18	168	42.28	288	9.34	720	0.56
336	17.21	336	47.54	576	9.84	1,440	0.63
504	7.79	504	27.31	864	6.54	2,160	0.54
672	7.39	672	24.81	1,152	8.80		
840	7.40	840	17.48	1,440	7.58		
1,008	6.95	1,008	23.73	1,728	10.31		

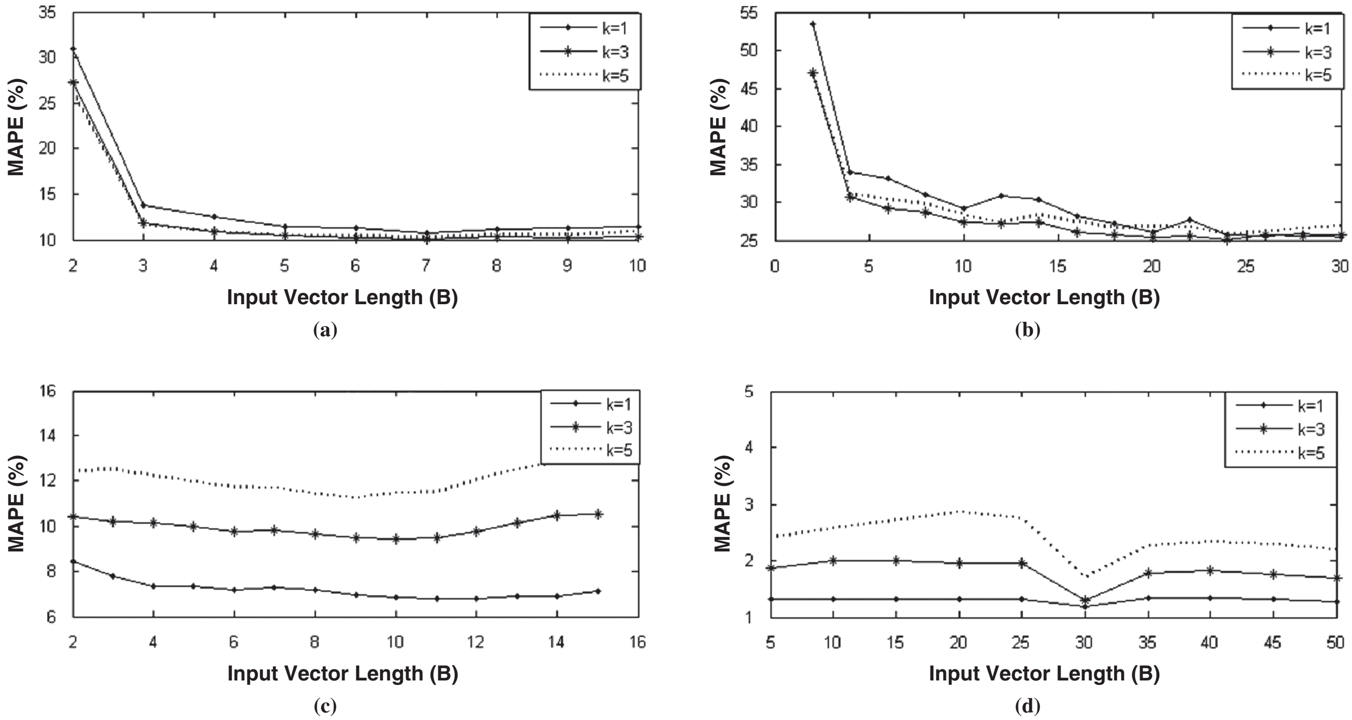

FIGURE 3 Performance of *k*-NN regarding input data vector length (*B*) and number of nearest neighbors (*k*): (*a*) I-90, (*b*) Peace Bridge, (*c*) Virginia, and (*d*) Beijing.

way to jointly determine the values of the two parameters is enumeration and testing. For simplicity, this paper reports only partial results, in which the value of one parameter is fixed and the other is varied and the impact of the variation on the MAPE is monitored. Lin et al. provide a detailed discussion of the selection of the other parameters for the SVR model (*29*).

With the value of the input vector initially fixed at 6 for all four data sets, the performance of SVR with respect to various training data sizes is shown in Table 3. The best training data size for the first two data sets appears to be around 600 data points. For the Virginia data, the lowest MAPE is attained when the training data size is equal to 300. For the Beijing data, there is a significant decrease in MAPE when the training data size is increased from 300 to 400, with the lowest MAPE achieved with data size equal to 700. This shows that for SVR, the appropriate moving training data set length is not really correlated to the Hurst exponent values, as was the case with the SARIMA model. This is because SVR does not consider the autocorrelation in the time dimension; instead, it considers a few support vectors to formulate the function.

Next, with the training data size fixed at either 100 or 500 data points, the length of the input vector was varied and the corresponding MAPEs were calculated. As shown in Figure 4, the vector length *D* affects the performance of SVR only for the Peace Bridge data set. For the I-90, Virginia, and Beijing data sets, the resultant MAPEs remain almost the same when the vector length varies from 2 to 10.

COMPARISON OF MODEL PERFORMANCE

The best models calibrated in the previous section were then applied to the four data sets. The results are shown in Table 4, along with the results from a naive prediction model, which basically uses the value of the observed volume at the current time step as the predicted value for the next step. The naive prediction model is used for benchmarking. The table also lists the values of the data diagnosis measures for each data set to provide a clear view of the relationship between data characteristics and model selection.

TABLE 3 Performance of SVR, Training Data Size

	MAPE (%) by Training Data Size									
Data Set	100	200	300	400	500	600	700	800	900	1,000
I-90	19.77	19.01	18.75	18.47	18.32	18.22	19.32	19.68	21.74	22.73
Peace Bridge	14.83	11.17	10.49	9.90	9.02	8.37	8.49	9.47	11.59	13.66
Virginia	37.58	29.7	13.58	13.80	17.22	17.11	17.33	18.20	19.03	20.41
Beijing	8.66	5.50	10.99	1.00	0.24	0.17	0.01	0.01	0.01	0.01

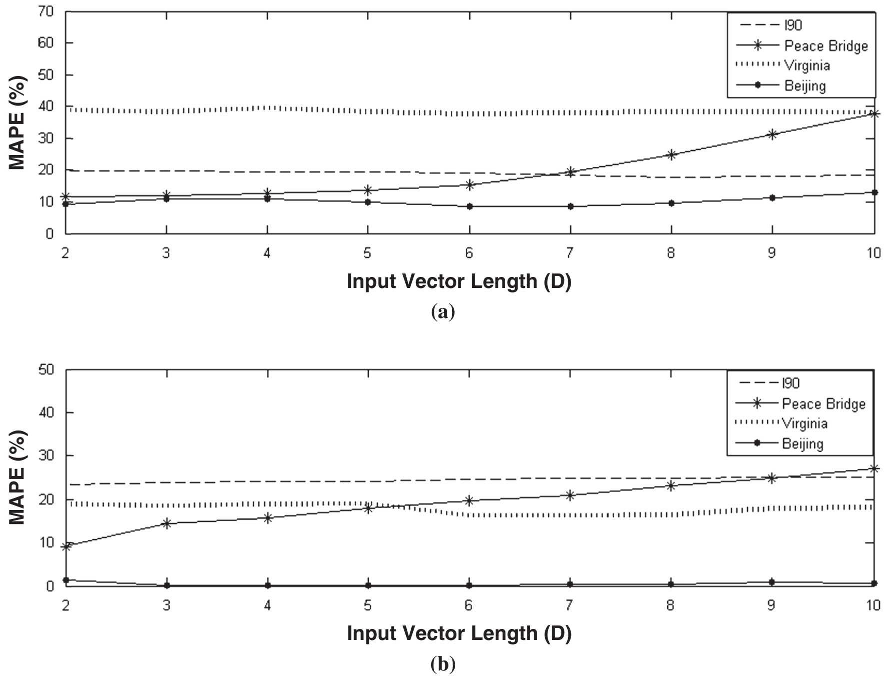

FIGURE 4 Performance of SVR regarding input data vector length *D:* (*a*) training data set length of 100 and (*b*) training data set length of 500.

Table 4 shows first, that all three models perform better than does the naive model. Second, for all four data sets, SARIMA performs slightly better than does the *k*-NN; however, the advantage of *k*-NN is its low computational cost and ease of implementation. Third, SARIMA and *k*-NN work much better for the I-90 and Virginia traffic volume data sets than does SVR, while SVR performs the best for the Peace Bridge and Beijing traffic volume data sets. This shows that SARIMA and *k*-NN are more appropriate for linear data sets and SVR is definitely a good choice for nonlinear data sets. Another observation is that for both linear data sets and nonlinear data sets, the larger the ApEn is, the higher the MAPE is.

CONCLUSIONS

Complexity, nonlinearity, and long-range dependency tests were used to assess the predictability of four traffic volume data sets. Three prediction models, SARIMA, *k*-NN, and SVR, were calibrated and tested for each of the data sets. The performances of the different prediction methods were then correlated to the results from the data diagnosis measures; this correlation provides some guidelines on how to choose the appropriate prediction method and set its parameters, given the statistical characteristics of a given data set.

Regarding prediction model choice, the following results were obtained:

1. SARIMA performed slightly better in all four data sets than did *k*-NN, but *k*-NN had a faster running speed;
2. SARIMA and *k*-NN were more appropriate for linear data sets, and SVR worked better than did SARIMA and *k*-NN for linear data sets; and
3. The larger the ApEn was, the higher the MAPE was.

Regarding setting the model parameters, the study found the following results:

1. SARIMA model. According to the Hurst exponent, the Peace Bridge data set had the weakest LRD, followed by the Virginia data set and then the I-90 data set. The best training data set length was 840 for Peace Bridge, 864 for Virginia,, and 1,008 for I-90. This

TABLE 4 Comparisons of Three Prediction Models for Four Data Sets

	Predictability			Performances [MAPE (%)]			
Data Set	Complexity (ApEn)	Nonlinearity (SurroData)	LRD (HurstEn)	SARIMA	*k*-NN	SVR	Naive Model
I-90	0.24	No	0.26	6.95	10.03	18.22	23.51
Peace Bridge	0.30	Yes	0.60	17.48	25.13	8.37	42.34
Virginia	0.21	No	0.69	6.54	6.81	13.58	15.23
Beijing	0.0049	Yes	0.91	0.54	1.18	0.01	14.59

NOTE: SurroData = surrogate data; HurstEn = Hurst exponent.

shows that for SARIMA, a weaker LRD indicated a smaller training data set length. For data sets like Beijing, with very small ApEn, the training data set length can also be set small.

2. *k*-NN model. The vector lengths corresponding to the lowest MAPEs were consistent with the results of the delay time and embedding dimension analysis. This was true for the I-90, Virginia, and Beijing data sets. For the Peace Bridge data set, the best vector length was 24 because of the data's strong seasonality.

3. *k*-NN. The number of nearest neighbors, *k,* should be set higher for data sets with higher ApEns.

4. SVR model. The training data set length did not appear to be sensitive to the Hurst exponent value. There also did not appear to be a strong relationship between the delay time and embedding dimension and the vector length, with the best value of the input vector length varying with changes in the training data set length.

For future research, the authors plan to undertake further testing on additional traffic volume data sets to ensure that the conclusions from this study may be generalized; this may include traffic volume data sets from arterials. Another suggested future research direction is to use the insights gained from this study to develop a decision support tool that could aid analysts in selecting the appropriate modeling paradigm for a given data set and in setting the values of the model's parameters.

REFERENCES

1. Williams, B. M., and L. A. Hoel. Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results. *Journal of Transportation Engineering,* Vol. 129, No. 6, 2003, pp. 664–672.
2. Smith, B. L., B. M. Williams, and R. K. Oswald. Comparison of Parametric and Nonparametric Models for Traffic Flow Forecasting. *Transportation Research Part C,* Vol. 10, No. 4, 2002, pp. 303–321.
3. Cools, M., E. Moons, and G. Wets. Investigating the Variability in Daily Traffic Counts Through Use of ARIMAX and SARIMAX Models: Assessing the Effect of Holidays on Two Site locations. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2136,* Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 57–66.
4. Smith, B. L., and M. J. Demetsky. Short-Term Traffic Flow Prediction: Neural Network Approach. In *Transportation Research Record 1453,* TRB, National Research Council, Washington, D.C., 1994, pp. 98–104.
5. Park, B., C. J. Messer, and T. Urbanik II. Short-Term Freeway Traffic Volume Forecasting Using Radial Basis Function Neural Network. In *Transportation Research Record 1651,* TRB, National Research Council, Washington, D.C., 1998, pp. 39–47.
6. Castro-Neto, M., Y.-S. Jeong, M.-K. Jeong, and L. D. Han. Online-SVR for Short-Term Traffic Flow Prediction Under Typical and Atypical Traffic Conditions. *Expert Systems with Applications,* Vol. 36, No. 3, 2009, pp. 6164–6173.
7. Hong, W.-C., and Y. Dong. Forecasting Urban Traffic Flow by SVR with Continuous ACO. *Applied Mathematical Modeling,* Vol. 35, No. 3, 2011, pp. 1282–1291.
8. Smith, B. L., and M. J. Demetsky. Traffic Flow Forecasting: Comparison of Modeling Approaches. *ASCE Journal of Transportation Engineering,* Vol. 123, No. 4, 1997, pp. 261–266.
9. Lam, W. H. K., Y. F. Tang, K. S. Chan, and M-L. Tam. Short-Term Hourly Traffic Forecasting Using Hong Kong Annual Traffic Census. *Transportation,* Vol. 33, No. 3, 2006, pp. 291–310.
10. Sun, H., H. X. Liu, H. Xiao, R. R. He, and B. Ran. Use of Local Linear Regression Model for Short-Term Traffic Forecasting. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1836,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 143–150.
11. Vlahogianni, E. I., M. G. Karlaftis, and J. C. Golias. Statistical Methods for Detecting Nonlinearity and Non-Stationarity in Univariate Short-Term Time-Series of Traffic Volume. *Transportation Research Part C,* Vol. 14, No. 5, 2006, pp. 351–367.
12. Shang, P., X. Li, and S. Kamae. Chaotic Analysis of Traffic Time Series. *Chaos, Solitons, and Fractals,* Vol. 25, No. 1, 2005, pp. 121–128.
13. Jayawardena, A. W., W. K. Li, and P. Xu. Neighborhood Selection for Local Modeling and Prediction of Hydrological Time Series. *Journal of Hydrology,* Vol. 258, No. 1–4, 2002, pp. 40–57.
14. Liu, S., Y. Zhang, S. Zhu, and Q. He. Study on Prediction of Chaotic Time Series Based on Phase Space Reconstruction. *Proc., 2011 International Conference on Informatics, Cybernetics, and Computer Engineering, November 19–20, 2011, Melbourne, Australia,* (J. Liangzhong, ed.), Springer, New York, 2011, pp. 87–97.
15. Fraser, A. M., and H. L. Swinney. Independent Coordinates for Strange Attractors from Mutual Information. *Physical Review A,* Vol. 33, No. 2, 1986, pp. 1134–1140.
16. Kennel, M. B., R. Brown, and H. D. I. Abarbanel. Determining Embedding Dimension for Phase-Space Reconstruction Using a Geometrical Construction. *Physical Review A,* Vol. 45, No. 6, 1992, pp. 3403–3411.
17. Ho, K. K., G. B. Moody, C. K. Peng, J. E. Mietus, M. G. Larson, D. Levy, and A. L. Goldberger. Predicting Survival in Heart Failure Case and Control Subjects by Use of Fully Automated Methods for Deriving Nonlinear and Conventional Indices of Heart Rate Dynamics. *Circulation,* Vol. 96, No. 3, 1997, pp. 842–848.
18. Schreiber, T., and A. Schmitz. Surrogate Time Series. *Physical D,* Vol. 142, No. 3–4, 2000, pp. 346–382.
19. Disks, C., J. C. van Houwelingen, F. Takens, and J. DeGoede. Reversibility as a Criterion for Discriminating Time Series. *Physical Letter A,* Vol. 201, No. 2–3, 1995, pp. 221–228.
20. Merkwirth, C., U. Parlitz, I. Wedekind, D. Engster, and W. Lauterborn. *TSTOOL User Manual,* Drittes Physikalisches Institut, University of Göttingen, Germany, 2009. http://www.physik3.gwdg.de/tstool/index.html.
21. Barbulescu, A., C. Serban, and C. Maftei. Evaluation of Hurst Exponent for Precipitation Time Series. *Proc., 14th WSEAS International Conference on Computers,* Corfu Island, Greece, World Scientific and Engineering Academy and Society, Athens, Greece, 2010, pp. 590–595.
22. Qian, B., and K. Rasheed. Hurst Exponent and Financial Market Predictability. *Proc., IASTED Conference on Financial Engineering and Applications,* Cambridge, Mass., ACTA Press, Calgary, Alberta, Canada, 2004, pp. 203–209.
23. Hurst, H. E. Long-Term Storage of Reservoirs: An Experimental Study. *Transactions of the American Society of Civil Engineers,* Vol. 116, 1951, pp. 770–799.
24. Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control.* John Wiley & Sons, Hoboken, N.J., 2008.
25. Vapnik, V. *The Nature of Statistical Learning Theory.* Springer, New York, 1995.
26. Kim, K.-J. Financial Time Series Forecasting Using Support Vector Machines. *Neurocomputing,* Vol. 55, No. 1–2, 2003, pp. 307–319.
27. Chang, C.-C., and C.-J. Lin. *LIBSVM: A Library for Support Vector Machines.* 2001. http://www.csie.ntu.edu.tw/cjlin/libsvm.
28. Wen, F., and Q. Wan. Time Delay Estimation Based on Mutual Information Estimation. *Proc., 2nd International Congression on Image and Signal Processing,* IEEE, New York, 2009, pp. 1–5.
29. Lin, L., Q. Wang, and A. W. Sadek. Multiple-Model Combined Forecasting Method for Online Prediction of Border Crossing Traffic at Peace Bridge. Presented at 91st Annual Meeting of the Transportation Research Board, Washington, D.C., 2012.

The Statistical Methods Committee peer-reviewed this paper.