Short-Term Traffic Flow Forecasting for Urban Roads Using Data-Driven Feature Selection Strategy and Bias-Corrected Random Forests

Jishun Ou, Jingxin Xia, Yao-Jan Wu, and Wenming Rao

Urban traffic flow forecasting is essential to proactive traffic control and management. Most existing forecasting methods depend on proper and reliable input features, for example, weather conditions and spatio-temporal lagged variables of traffic flow. However, the feature selection process is often done manually without comprehensive evaluation and leads to inaccurate results. For that challenge, this paper presents an approach combining the bias-corrected random forests algorithm with a data-driven feature selection strategy for short-term urban traffic flow forecasting. First, several input features were extracted from traffic flow time series data. Then the importance of these features was quantified with the permutation importance measure. Next, a data-driven feature selection strategy was introduced to identify the most important features. Finally, the forecasting model was built on the bias-corrected random forests algorithm and the selected features. The proposed approach was validated with data collected from three types of urban roads (expressway, major arterial, and minor arterial) in Kunshan City, China. The proposed approach was also compared with 10 existing approaches to verify its effectiveness. The results of the validation and comparison show that even without further model tuning, the proposed approach achieves the lowest average mean absolute error and root mean square error on six stations while it achieves the second-best average performance in mean absolute percentage error. Meanwhile, the training efficiency is improved compared with the original random forests method owing to the use of the feature selection strategy.

One overarching objective of intelligent transportation systems is to improve the operational efficiency of transportation systems in urban areas. To achieve this goal, a spectrum of advanced and diverse techniques is now being deployed to gather, analyze, manage, and disseminate information across transportation networks. Among these techniques, traffic flow forecasting has consistently attracted a great deal of attention from researchers around the world, especially in the past two decades. It is widely recognized that traffic flow forecasting can help reduce travel times and costs for commuters, cut environmental pollution, and support the efforts of traffic authorities to engage in proactive traffic management (*1*).

Traffic demand forecasting, which relies on accurate and robust traffic volume forecasting, has been identified as one of the most important yet challenging issues for developing advanced intelligent transportation systems applications (*2*). Urban road networks present more complicated traffic flow characteristics than highways or freeways, with frequent disturbances resulting from communication disruptions or traffic control strategies, and thus pose a serious challenge for effective traffic forecasting. The main objective of this study is therefore to develop a traffic forecasting approach that can effectively and efficiently predict traffic volume for a dense urban road network.

During the past few decades, a wide variety of approaches have been proposed for traffic flow forecasting. These include time series models such as the autoregressive integrated moving average (ARIMA) (*3–5*) and its variant, seasonal ARIMA (*6–9*), historic average (*10*), Kalman filter (*11, 12*), nonparametric regression (*13–17*), support vector regression (*18–20*), and neural networks (*21–26*), among others. Although all of these methods have been shown to be both feasible and effective, they each suffer from drawbacks that limit their performance under various conditions. For instance, the historic average model finds it difficult to observe rapid variations in flow in urban areas (*27*), and although ARIMA and Kalman filter models can forecast stable traffic flow timely and accurately, neither can model nonlinear relationships resulting from a structural change in a traffic flow series effectively because they are linear models (*28*). The nonparametric regression model is intuitive in its formulation and easy to implement, but its predictive ability depends heavily on the capacity of the historical database available (*15*), and while the neural network model is capable of describing the indeterministic and complex nonlinearities in traffic flow, it struggles to provide meaningful results that can be easily understood by decision makers as it belongs to the family of "black-box" models (*29*). Moreover, for some types of neural network models, rich experience and expertise are needed to set proper parameters. The support vector regression model has a solid theoretical foundation based on statistical learning theory and the structural risk minimization principle, but still requires significant expertise and time for calibration if satisfactory results are to be obtained, and if the number of predictor variables is large, the training process will be very time consuming (*30*). Vlahogianni et al. provided a comprehensive literature review on this topic (*31*).

As pointed out by Smith et al., there is no single model that consistently has an absolute advantage over a range of different conditions

J. Ou, J. Xia, and W. Rao, Intelligent Transportation System Research Center, Southeast University, No. 35 Jinxianghe Road, Nanjing 210096, China. Y.-J. Wu, Department of Civil Engineering and Engineering Mechanics, College of Engineering, University of Arizona, Room 324F, 1209 East 2nd Street, Tucson, AZ 85721. Corresponding author: J. Xia, xiajingxin@seu.edu.cn.

Transportation Research Record: Journal of the Transportation Research Board, No. 2645, 2017, pp. 157–167.
http://dx.doi.org/10.3141/2645-17

because traffic flow generally exhibits complex spatiotemporal behaviors characterized by irregular randomness, making it very difficult for a single forecasting model to capture all of these perturbed patterns (*14*). This observation is true especially for urban environments. In view of this, some researchers have sought to build combined models that take advantage of the strengths of multiple forecasting methods. For example, Zhang and Ye presented a short-term traffic flow forecasting approach based on fuzzy logic system methodology that combines the strengths of several different models under varying traffic conditions, showing that the combined model can indeed produce more accurate and stable forecasts than individual forecasting models such as ARIMA and Kalman filters (*32*). Tan et al. proposed an aggregation model based on four different algorithms—moving average, exponential smoothing, ARIMA, and neural networks—arguing that this type of aggregation strategy can offer substantial benefits in improving operational forecasting (*33*). Similarly, Wang et al. used a Bayesian combination method to improve the forecasts of single predictors by linearly incorporating ARIMA, Kalman filter, and a backpropagation neural network into their combination framework and demonstrating the effectiveness of this approach with a numerical simulation (*34*).

Despite the fact that a wide variety of methods have been proposed, gradually evolving from linear forms to nonlinear forms and from univariate modeling to multivariate modeling, most of the relevant studies focus on designing new forecasting algorithms or improving existing ones. The performance of the models depends not only on the algorithm but also on the input features. The models could give inaccurate or unstable forecasts when improper input features are used (*35–37*). Since traffic flow always exhibits a variety of characteristics and is influenced by multiple factors, a number of relevant features can be extracted from traffic flow time series data, such as temporal lagged features used to represent temporal correlations; multivariate lagged features used to represent dynamic interactions between flow, speed, and occupancy; or spatial lagged features used to represent spatial correlations (*29*). Not all of the extracted features are useful for improving forecasting performance owing to the existence of irrelevant or redundant features in the candidate feature pool. Several researchers have noticed this problem and tried to solve it by adopting some prior feature selection strategies (*19–21*). For example, Wei and Liu (*19*) used the statistical autocorrelation function to determine the suitable temporal lagged features for the support vector regression model, while the spectral and cross-spectral analysis techniques were used by Stathopoulos and Karlaftis to analyze urban traffic flow characteristics before training a neural network model (*21*). A common disadvantage of these two methods is that only the spatiotemporal lagged features were extracted and considered, which implies there is still room for improvement if more comprehensive features are considered.

Based on the above analysis, this paper presents an approach combining the bias-corrected random forests (BCRF) algorithm with a data-driven feature selection strategy for short-term urban traffic flow forecasting. Adopting feature extraction and feature selection allows identification of the most important features that can be used to further enhance forecasting performance. Using the BCRF algorithm reduces the forecasting bias stemming from the original RF algorithm (*38*). Although these general methods and tools have already been developed, their application in traffic flow forecasting and their integration into the proposed framework of this approach are new.

METHOD

Framework of Proposed Approach

The overall framework of the proposed approach is shown in Figure 1. In general, the framework consists of three critical parts: (*a*) extract representative features from traffic time series to effectively represent the intricate characteristics of dynamic and stochastic traffic processes; (*b*) use data-driven feature selection strategy to select and distinguish the most useful features from the candidate feature pool of extracted features to improve the accuracy of the forecasting model further, or at least maintain the level of accuracy achieved without any selection; and (*c*) train a BCRF model with the selected features to produce accurate and robust forecasts. In this study, the focus is on using this framework to implement short-term traffic volume forecasting for urban roads. The framework can be applied in other traffic flow forecasting applications (e.g., short-term speed forecasting) and other road types (e.g., freeways and rural highways). It can also be incorporated into a number of different forecasting algorithms and feature selection strategies.

Feature Extraction

For the analysis of complex data, feature selection is critical. When short-term traffic flow is forecast, the time series of traffic volume, speed, and occupancy collected from different locations usually exhibits a number of different characteristics, so extracting representative features to reflect these characteristics becomes a significant preliminary step. Ma et al. summarized the characteristics of a time series of traffic flow in relation to the following properties: multiseasonality, nonstationarity, temporal and spatial correlations, and dynamics between traffic variables that lead to the interaction of multivariate traffic time series (*29*). Four categories of features corresponding to these characteristics are considered in this paper, as follows:

1. Temporal lagged features (F_T). Temporal correlations in the time series of traffic flow can be described by time-lagged features. For example, the first-order autocorrelation is the temporal correlation between a current entry in the time series and its previous entry. Likewise, multiple time-lagged features can be used to describe a series of intricate temporal correlations in a traffic flow time series. In this study, for the time series of traffic volume, the first 12 lags $Q_{\text{lag1}} \sim Q_{\text{lag12}}$ were extracted as the temporal lagged features. Theoretically, the order of lags can be set to an arbitrarily large value. However, the bigger the order of lags setup is, the more features will be extracted. Correspondingly, the time spent on the process of feature selection will be longer.
2. Nominal time features (F_P). The time series of a traffic flow may change seasonally or fluctuate regularly. For instance, traffic volume is likely to follow a daily cycle, a monthly cycle, and even an annual cycle. These regular fluctuations are referred to as multiseasonality and can be represented by nominal time features. Six categories of nominal time features were considered in the later experiments, including IsNight, MinuteOfHour, HourOfDay, DayOfWeek, WeekOfMonth, and MonthOfYear.
3. Interaction features (F_D). From the viewpoint of macroscopic traffic flow theory, traffic volume, speed, and density are closely related to one another. The corresponding association between them is often referred to as the fundamental diagram. To capture the dynamics between traffic variables, it is necessary to incorporate

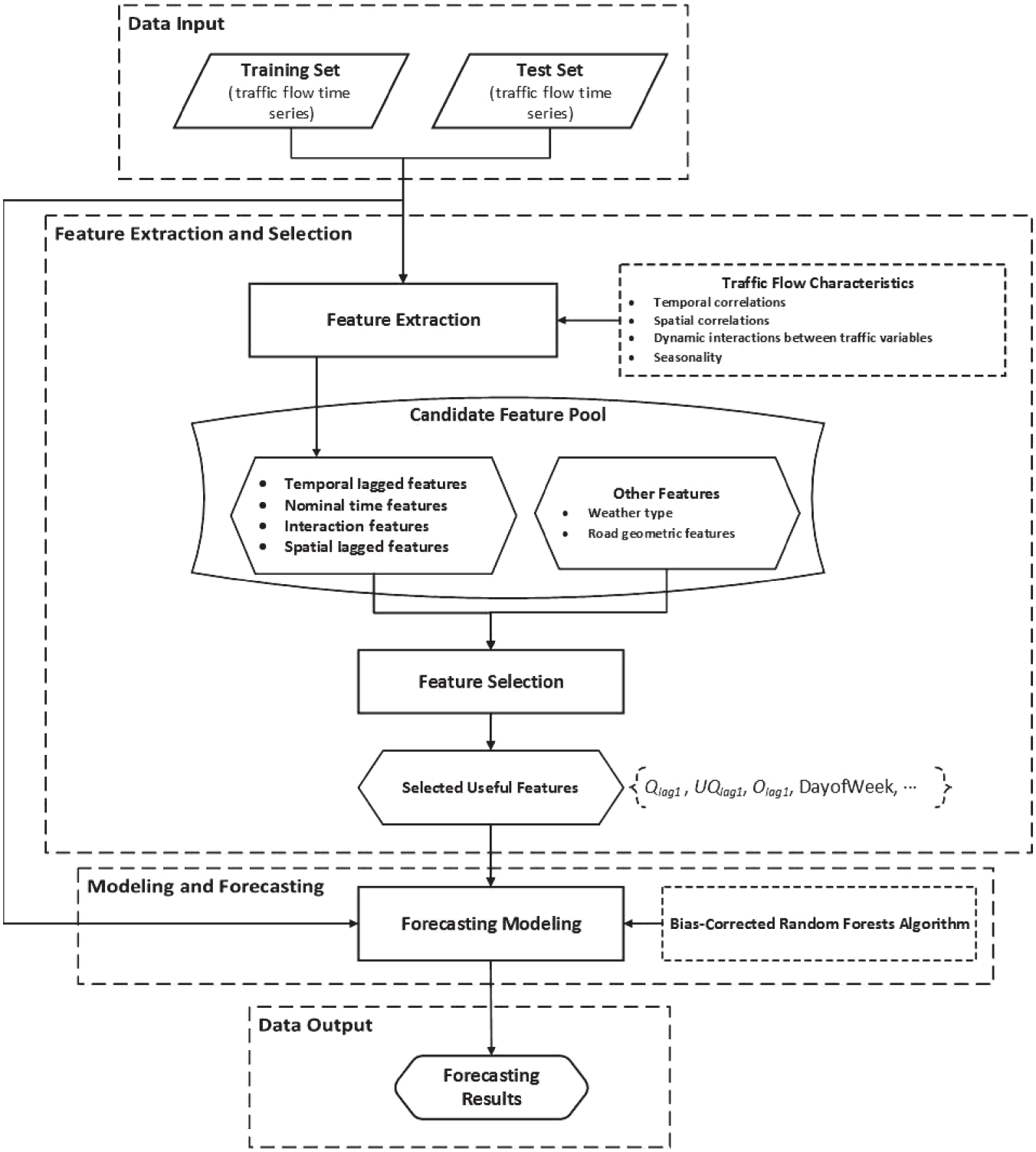

FIGURE 1 Overall framework of proposed approach.

the time-lagged features of multivariate time series during the modeling procedure. These multivariate time-lagged features are denoted as interaction features. In this study, the time series of speed and occupancy was presumed to have potential interaction with the time series of volume. Correspondingly, the lagged speed features $V_{lag1} \sim V_{lag12}$ and the lagged occupancy speed features $O_{lag1} \sim O_{lag12}$ were extracted to represent the interaction features.

4. Spatial lagged features (F_S). In addition to temporal correlations, spatial correlations represent another kind of underlying characteristic in traffic flow time series. It is well known that the traffic flow upstream of a given location is likely to affect the downstream traffic, so these spatial correlations should also be embedded in the forecasting model. In this situation, time-lagged features of multivariate time series at neighboring sites are considered, including the upstream lagged volume features $UQ_{lag1} \sim UQ_{lag12}$, the upstream lagged speed features $UV_{lag1} \sim UV_{lag12}$, and the upstream lagged occupancy features $UO_{lag1} \sim UO_{lag12}$.

Other features can also be involved, including weather descriptions, such as temperature, humidity, precipitation, and wind power, as well as features related to static road conditions such as the number of lanes. In short, more features can be poured into the candidate feature pool to strengthen the predictive power of the forecasting model.

Feature Selection

A wide spectrum of candidate features can be achieved through feature extraction. However, not all extracted features are useful for improving forecasting accuracy since there may be some irrelevant or redundant features in the extracted features. When these features are removed, the forecasting accuracy could be further improved, or could at least maintain the accuracy (*35–37*). It thus makes sense to carry out feature selection after the feature extraction process.

A vast number of feature selection strategies have been proposed for applications in different fields. A recent review by Chandrashekar and Sahin provides a comprehensive overview of popular feature selection strategies (*36*). Generally, these strategies can be divided into the following three types according to how they combine the selection algorithm and the model building: (*a*) filter strategies, (*b*) wrapper strategies, and (*c*) embedded strategies. For filter strategies, the feature importance measure is not related to a given model design method, whereas in wrapper strategies the model forecasting performance is included in the feature score calculation procedure. Embedded strategies combine feature selection and model estimation more closely; in a sense, the model training process integrates the selection algorithm. Genuer et al. proposed an embedded feature selection strategy and demonstrated its effectiveness through a series of experiments (*37*). Their strategy was chosen for the current study for two reasons: (*a*) it is based entirely on random forests (RF), which is also used to build the forecasting model in this paper so it can be seamlessly incorporated into the overall framework, and (*b*) it is effective and can be used in classification and regression problems, which is consistent with the objective of this paper, in which traffic volume forecasting is viewed as a kind of regression problem.

Since the RF algorithm provides two useful feature importance measures, that is, the permutation importance measure and the Gini impurity measure, with which to select informative features, the next subsection will discuss the details of the relevant importance measures, after which a two-step feature selection strategy will be represented.

Permutation Importance Measure in RF

RF is a nonparametric method that builds an ensemble model of classification and regression trees (CART) (*39*) from random subsets of features and bagged samples of training data [see details in Breiman (*38*)]. A powerful property of RF lies in its ability to measure feature importance during the training process without invoking extra computing steps. Two measures are provided by regression RF: (*a*) a permutation importance measure and (*b*) a Gini impurity measure. In the feature selection strategy used in this paper, only the permutation importance measure is involved; thus, the Gini impurity measure will not be discussed here (*37*).

The permutation importance is defined as the mean decrease in accuracy that occurs because of the out-of-bag observations. Since each tree in RF is grown from a bootstrapped sample, on average, about one-third of the observations in the data set will not be used to grow the tree. These observations are considered the out-of-bag (OOB) observations for that tree and form a natural test set for each tree. For any feature X^m ($1 \le m \le M$) in a given feature set X, M is the number of features in X, consider OOB_k as the associated OOB sample for each tree T_k ($1 \le k \le K$) in RF, $ErrOOB_k$ as the error (mean square error for regression) of T_k on this OOB_k sample, and K is the number of trees in RF. Now, randomly permute the values of X^m in OOB_k to obtain a perturbed sample, denoted by $\widetilde{OOB}_k^m$, and compute the corresponding error $Err\widetilde{OOB}_k^m$ for tree T_k. The permutation importance of X^m is therefore equal to

$$PI_m = \frac{1}{K}\sum_{k=1}^{K}\left(Err\widetilde{OOB}_k^m - \widetilde{OOB}_k^m\right) \tag{1}$$

For a fixed number of trees, a feature with a larger permutation importance (PI) relative to other features indicates that it is more informative in the forecasting task.

Two-Step Feature Selection Strategy

Step 1. Preliminary ranking and elimination:

– Rank the features by sorting the PI indicator (averaged from the 50 runs) in descending order.

– Remove those features with an average PI less than the minimum forecasting value given by a CART model fitting the curve of the corresponding standard deviations of PI. Denote by M_r the number of remaining features.

Step 2. Feature selection for two distinctive objectives:

– For interpretation. Construct the nested collection of RF models involving the k first features kept in Step 1, for $k = 1$ to M_r, and select the features involved in the model leading to the smallest OOB error (averaged on 50 runs and using default RF parameters).

– For prediction. Starting from the ordered features retained for interpretation, construct an ascending sequence of RF models by invoking and testing the features stepwise; a feature is selected only if the error gain exceeds a threshold determined by the average variation obtained by adding noisy features, defined as follows:

$$\frac{1}{p_{elim} - p_{interp}}\sum_{i=p_{elim}}^{p_{interp}}\left|ErrOOB(i+1) - ErrOOB(i)\right| \tag{2}$$

where p_{elim} is the number of features removed in Step 1 (noisy features), p_{interp} is the number of features selected for interpretation in

Step 2, and ErrOOB(j) is the OOB error of the RF built using the j most important features.

BCRF Model for Forecasting

Although experiments [e.g., Caruana et al. (*30*) and Genuer et al. (*37*)] have demonstrated that RF has excellent performance in comparison with typical predictive algorithms, the predicted value of RF can be biased in that large and small values in the leaf node samples are often either underestimated or overestimated (*40*). To reduce bias in general forecasting problems, several bias-correction strategies have been proposed (*40, 41*). Two simple yet effective strategies are introduced here to correct for the bias in RF: the first uses the median instead of the mean as the aggregation rule for integrating individual tree forecasts, and the other uses the residual term of the first-level tree models as the response feature to train the second-level tree models for bias forecasting, and then combines the corresponding outputs of these two-level models to obtain the corrected forecasts. The algorithm for this BCRF model is described as follows:

Step 1. Draw a bootstrap sample TR_k from the original training data set TR, that is, sampling with replacement using a bagging technique (*42*).

Step 2. Grow an unpruned regression tree T_k for each TR_k. At each node t, rather than choosing the best split of all features, randomly sample m of the features from the total M features, and compute the best split from the chosen features. The split is determined by the mean decrease in impurity, defined as follows:

$$\mathrm{IM}_t = \frac{\sum_{x_i \in t} \left(Y_i - \bar{Y}_t\right)^2}{N_t} \tag{3}$$

where

N_t = number of training samples,
x_i = predictor feature used to split node,
Y_i = response feature at node t, and
$\bar{Y}_t$ = mean of all Y_i.

If t is a leaf node, $\bar{Y}_t$ is assigned as the forecasting value of this node.

Step 3. Repeat Step 1 and Step 2 K times, thus growing K trees.

Step 4. Given a new input x, the forecasts for the K trees are computed separately, denoted as $\widetilde{f_1(x)}, \widetilde{f_2(x)}, \ldots, \widetilde{f_k(x)}, \ldots, \widetilde{f_K(x)}$. Use the forecasting biases of these trees, that is, $f(x) - \widetilde{f_1(x)}, f(x) - \widetilde{f_2(x)}, \ldots, f(x) - \widetilde{f_k(x)}, \ldots, f(x) - \widetilde{f_K(x)}$ to replace the values of the response feature in the original training data set, where $f(x)$ is the true value of the response feature. The newly created data set is then used to train the second-level tree models. The bias-corrected values of these K trees can now be calculated as the forecasting value of the first-level model plus the forecasting value of the second-level model. Denote $\widetilde{f'_1(x)}, \widetilde{f'_2(x)}, \ldots, \widetilde{f'_k(x)}, \ldots, \widetilde{f'_K(x)}$ as the bias-corrected values of the K trees.

Step 5. Aggregate the bias-corrected forecasts of the K trees with a median rule to obtain the final forecast of x (one-step-ahead forecast):

$$\widetilde{f(x)} = \mathrm{median}\left(\widetilde{f'_k(x)}\right)(1 \le k \le K) \tag{4}$$

DATA DESCRIPTION

Traffic flow data for three types of urban roads (expressway, major arterial, and minor arterial) in Kunshan City, China, were used to validate the performance of the proposed approach. The data were collected through microwave detectors installed across all lanes of traffic. Of these, six stations were chosen to test model performance in the later experimental analysis. The relevant traffic measurements collected were volume, speed, and occupancy, aggregated at 5-min intervals for 60 days, from May 2 to June 30, 2016. Data for 53 days (15,264 data records for each station), from May 2 to June 23, were set aside as the training set to construct the forecasting models, and the other 7 days (2016 data records for each station), from June 24 to June 30, were used as the test set to validate the models' performance. Here, the data for the first 53 days were used as the training set and the data for the other 7 days as the test set because it was assumed it was possible to obtain a reliable forecast based on the partitioned training set. Meanwhile, the 7-day test set covers different traffic patterns in volume time series and thus can be used to evaluate the performance of different forecasting models. Several preprocessing operations were conducted to ensure good data quality, including duplicate removal and noise filtering. Table 1 describes the selected stations and their locations.

IDENTIFICATION OF INFORMATIVE FEATURES

One advantage of the proposed approach lies in its ability to identify informative features from a candidate feature pool filled with extracted features from traffic flow series. Table 2 illustrates the selected ratio (the number of total features divided by the number of selected features) for each category of extracted features. As mentioned earlier, four feature categories can be extracted, namely, F_T, F_D, F_S, and F_P, representing four kinds of traffic flow characteristics. As indicated in the table, no more than two-fifths of the total features were preserved after the feature selection process, confirming the existence of useless (irrelevant or redundant) features.

TABLE 1 Station Locations

Road Type	Road Name	Direction	Station ID	Number of Lanes	Speed Limit (km/h)
Expressway	Jiangpu Road	Southbound	Ex-1	2	80
		Northbound	Ex-2	2	80
Major arterial	Qianjin Road	Westbound	MaA-1	2	65
		Westbound	MaA-2	2	65
Minor arterial	Zhenchuan Road	Westbound	MiA-1	2	55
		Eastbound	MiA-2	2	55

TABLE 2 Selected Ratios of Extracted Features

Candidate Feature Pool	Ex-1	Ex-2	MaA-1	MaA-2	MiA-1	MiA-2
F_T	5/12	8/12	7/12	7/12	8/12	8/12
F_D	6/24	1/24	1/24	3/24	9/24	4/24
F_S	10/36	6/36	4/36	7/36	6/36	14/36
F_P	2/6	2/6	3/6	3/6	2/6	3/6
All	23/78	17/78	15/78	20/78	25/78	29/78

Figure 2 presents the quantification of the selected features computed by the permutation importance measure in RF. As can be seen from the figure, the selected features and their importance quantifications are distinct from stations, implying that there are diverse traffic patterns in the traffic flow for different road stations. However, the top three of the most important features are identical if the stations are grouped by road types, with the exception of one minor arterial group. In that group, stations MiA-1 and MiA-2 have the same top two features. In addition, for the three traffic measures considered, only the speed measure is rarely chosen in the feature selection process, reflecting the fact that the speed of traffic flow in Kunshan City does not fluctuate as violently as the flow and occupancy measures and is thus not a significant factor in traffic volume forecasting (*43*). Of the nominal time features, the most informative is HourOfDay, followed by MinuteOfHour. The indication is that the predicted flow series has an underlying hourly pattern that could be captured to enhance forecasting ability. Another useful feature is IsNight, which is selected on four stations. Although this feature has seldom been discussed in previous work, it can substantially contribute to forecasting accuracy. In contrast, another frequently mentioned feature, DayOfWeek, nearly disappears. The results of the feature selection reveal another interesting fact: the importance of the lagged features is not closely associated with their order sequence. For instance, for station MaA-1, Q_{lag1} is more important than Q_{lag3} but less informative than Q_{lag2}. This finding may imply that simply choosing the first few lagged variables as input features is not a wise decision in traffic forecasting modeling.

PERFORMANCE EVALUATION

To evaluate the effectiveness of the proposed approach, three performance measures are used: mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE), all of which are widely used to evaluate traffic forecasting performance (*44*). In addition, two categories of representative forecasting models were implemented: (*a*) six univariate models, NAÏVE, ARIMA, NNR_{uv}, $CART_{uv}$, $BCCART_{uv}$, and RF_{uv}, and (*b*) six multivariate models, including NNR_{mv}, $CART_{mv}$, $BCCART_{mv}$, RF_{mv}, $FSRF_{mv}$, and $FSBCRF_{mv}$.

Of these models, NAÏVE is the most trivial forecasting model (*45*). All forecasts of this model are simply set to be the value of the last observation in this study, and it is effectively used here as a baseline. For ARIMA, an automatic version was implemented with the "forecast" R package, which estimates the best parameters, that is, *p, d, q,* through the corrected Akaike information criterion (*46*). Recent research has highlighted the potential utility of nearest neighbor regression (NNR) models that combine forecast accuracy with short computation time, so the univariate version NNR_{uv} and the multivariate version NNR_{mv} are included here (*15–17*). The number of nearest neighbors *k* is determined by testing *k* from 1 to 10 and choosing the *k* that leads to the smallest MAPE. The "FNN" package was used to implement these two models (*47*).

CART is the component forecasting model for RF and can be used for classification and regression problems, so the univariate version $CART_{uv}$ and multivariate version $CART_{mv}$ are included using the "rpart" package (*48*). Since CART could be biased as the RF algorithm, two bias-corrected versions of CART ($BCCART_{uv}$ and $BCCART_{mv}$) models were also considered. Finally, two versions of the original RF, that is, the univariate version RF_{uv} and the multivariate version RF_{mv}; the improved RF model $FSRF_{mv}$, which adopts the feature selection strategy; and the proposed model $FSBCRF_{mv}$, which employs feature selection and bias correction, were also included in the test framework. These combined models were implemented on the basis of the "randomForests" package in R language (*49*). The number of component tree models was set at 150 by considering a reasonable trade-off between accuracy and efficiency.

Accuracy Comparisons

Table 3 shows the forecasting performance of the competing models on six data sets when the MAE, RMSE, and MAPE measures are used. For clarity, the best performance in each column is highlighted in bold.

As can be seen from the table, $FSBCRF_{mv}$ achieved the best or near-best performance in most cases. For the MAE and RMSE measures, $FSRF_{mv}$ and $FSBCRF_{mv}$ performed better than the other models, while NNR_{uv}, NNR_{mv}, and $FSBCRF_{mv}$ had an advantage when the MAPE measure was used. For NNR, CART, BCCART, and RF, two versions of each were implemented, namely, the univariate and multivariate versions. The pairwise comparative analysis demonstrates that the multivariate version was almost always more accurate than the corresponding univariate version, confirming that modeling traffic flow, which is a complex stochastic and dynamic process, is a challenging task that requires considering and integrating a number of different traffic characteristics (*29*).

However, in some situations, several multivariate models performed relatively poorly compared with the univariate version. For instance, on data sets Ex-2 and MiA-2, the univariate version of NNR and CART achieved a higher accuracy than the multivariate version in regard to MAE and RMSE performance measures. The same situation occurs for the NNR method on data sets Ex-1 and MiA-1 assessed by the MAPE performance measure. These special cases may imply that not all input features can be harnessed to improve the forecasting performance, but instead may actually degrade the forecasting ability of the model in some circumstances. However, the RF model seems to be less affected by this issue. The MAE, RMSE, and

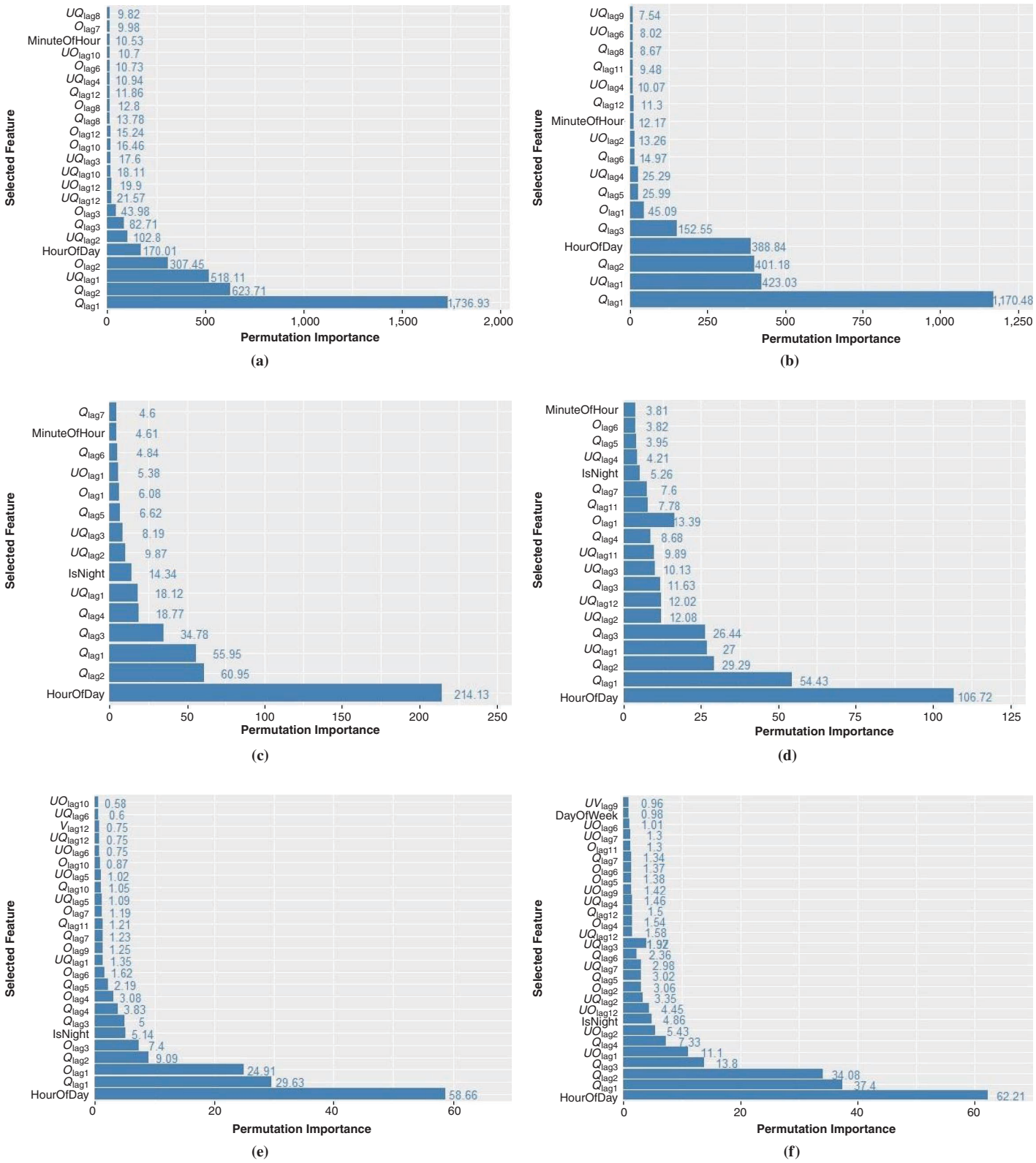

FIGURE 2 Permutation importance of selected features: (*a*) Ex-1, (*b*) Ex-2, (*c*) MaA-1, (*d*) MaA-2, (*e*) MiA-1, and (*f*) MiA-2.

TABLE 3 Forecasting Performance of Competing Models

Model	Ex-1			Ex-2			MaA-1		
	MAE (veh/5 min)	RMSE (veh/5 min)	MAPE (%)	MAE (veh/5 min)	RMSE (veh/5 min)	MAPE (%)	MAE (veh/5 min)	RMSE (veh/5 min)	MAPE (%)
NAÏVE	12.79	18.02	22.82	11.11	16.42	27.60	9.81	13.05	23.66
ARIMA	11.36	16.07	20.10	10.66	15.85	30.95	7.47	9.90	18.47
NNR_{uv}	10.26	14.48	**16.21**	10.74	16.63	25.61	7.41	9.85	17.05
$CART_{uv}$	16.48	22.13	41.20	14.10	19.36	50.18	9.09	12.13	23.23
$BCCART_{uv}$	12.77	17.97	23.81	11.42	16.73	30.69	8.25	10.92	21.52
RF_{uv}	10.23	14.38	18.56	9.43	14.24	27.26	7.17	9.58	17.90
NNR_{mv}	9.98	14.13	16.56	11.04	16.79	23.09	7.35	9.84	**16.46**
$CART_{mv}$	16.50	22.08	41.21	14.26	19.62	50.23	8.21	10.88	21.39
$BCCART_{mv}$	12.26	17.79	20.99	11.39	16.81	30.81	7.66	10.16	18.99
RF_{mv}	9.95	14.21	17.82	9.31	13.89	28.12	6.97	9.26	17.67
$FSRF_{mv}$	**9.69**	**13.72**	17.08	8.98	**13.61**	25.58	6.81	9.05	17.00
$FSBCRF_{mv}$	9.73	13.73	16.79	**8.86**	13.66	**21.78**	**6.79**	**9.04**	16.78

MAPE of the multivariate version of the RF model are consistently lower than those achieved by the univariate version. This finding is likely the result of the fact that RF is a combined forecasting method that takes advantage of the strengths of multiple single forecasting models to partly compensate for the effect of the useless features on forecasting accuracy. Nevertheless, one can see from the table that the improved version of RF_{uv}, namely, $FSRF_{mv}$, which uses a feature selection strategy, advances the forecasting performance further. As expected, the use of a bias correction mechanism also boosts the performance of the $FSBCRF_{mv}$ model.

Table 4 gives the average forecasting performance on six stations. Comprehensively, the $FSBCRF_{mv}$ model achieves the lowest average MAE and RMSE on six stations, while it achieves the second-best average performance in relation to the MAPE. What is somewhat surprising is that the CART model, whether the univariate version or the multivariate version, performed very poorly and was even less accurate than the NAÏVE model. Benefitting from its bias correction strategy, the BCCART model achieved better results than the CART model. The clear difference in performance between CART and RF indicates that the combined forecasting strategy can significantly improve the accuracy of the constructed models (*30*). Finally, the parametric ARIMA model failed to produce better results than the nonparametric NNR model, probably because of the existence of complex and dynamic nonlinear relationships in urban traffic flows.

Efficiency Comparisons

In addition to accuracy, efficiency is a key factor that must be considered when forecasting models are evaluated. Table 5 shows the training time and the forecasting time of each of the 12 competing models. The training time is the time required to estimate each model with the use of the training data set, while the forecasting time is the time required to calculate the forecasting values of all test samples. The computation time for each model is calculated and recorded by the same computer, which is equipped with a 3.2 GHz 4-core CPU and an 8 GB RAM. All of the forecasting models are implemented in the R language. As can be seen from the table, the combined models, that is, RF_{uv}, RF_{mv}, $FSRF_{mv}$, and $FSBCRF_{mv}$, required much more computing time than the single models since multiple replications need to be integrated into the final decision framework. However, there was little difference between the combined models and the single models when test samples were being predicted. Since NNR belongs to the family of lazy-learning algorithms, most of its computing time is spent seeking nearest neighbors in the historical database during the forecasting period, so it is not surprising that it had the longest forecasting times. The multivariate models will always be costlier than the corresponding univariate version during the training period, but not for forecasting. Both $FSRF_{mv}$ and $FSBCRF_{mv}$ required less training time than RF_{mv}. This observation demonstrates another advantage of the feature selection strategy, as it significantly reduced the training time of the forecasting models. However, the $FSBCRF_{mv}$ model still required a longer training time because of the introduction of the dual mechanism, which included feature selection and bias correction. However, since the model can be implemented with offline training, the deployment and maintenance of this model should not be a significant concern for practical applications.

TABLE 4 Average Forecasting Performance of Competing Models

Model	Average MAE (veh/5 min)	Average RMSE (veh/5 min)	Average MAPE (%)
NAÏVE	9.33	12.87	25.47
ARIMA	7.86	10.89	22.10
NNR_{uv}	7.68	10.75	19.48
$CART_{uv}$	10.12	13.53	32.01
$BCCART_{uv}$	8.66	11.90	24.45
RF_{uv}	7.36	10.21	21.29
NNR_{mv}	7.57	10.58	18.88
$CART_{mv}$	9.97	13.30	31.55
$BCCART_{mv}$	8.36	11.60	22.96
RF_{mv}	7.17	9.95	20.97
$FSRF_{mv}$	6.97	9.70	19.80
$FSBCRF_{mv}$	6.96	9.70	18.95

MaA-2			MiA-1			MiA-2		
MAE (veh/5 min)	RMSE (veh/5 min)	MAPE (%)	MAE (veh/5 min)	RMSE (veh/5 min)	MAPE (%)	MAE (veh/5 min)	RMSE (veh/5 min)	MAPE (%)
9.16	12.37	22.83	6.05	8.12	30.69	7.07	9.26	25.20
7.13	9.64	17.84	4.84	6.47	24.86	5.69	7.41	20.40
7.05	9.50	17.60	4.90	6.63	**21.63**	5.71	7.41	18.75
8.66	11.42	22.42	5.76	7.63	30.14	6.62	8.50	24.86
7.91	10.53	20.41	5.48	7.34	27.79	6.13	7.93	22.46
6.93	9.32	18.11	4.81	6.48	25.08	5.60	7.25	20.84
6.75	9.15	16.52	4.84	6.47	22.67	5.46	7.08	**18.42**
8.50	11.18	22.08	5.70	7.53	29.52	6.62	8.50	24.86
7.62	10.12	20.07	5.16	6.88	24.30	6.04	7.86	22.60
6.64	8.92	17.52	4.69	6.26	24.46	5.45	7.13	20.25
6.47	**8.77**	16.60	**4.61**	**6.14**	23.47	**5.27**	**6.93**	19.06
6.44	**8.77**	**16.28**	4.63	6.17	23.20	5.28	6.94	18.89

CONCLUSION

In this paper, an effective approach for urban traffic flow forecasting is presented in the context of multivariate modeling. The proposed approach can be divided into three parts. First, a number of features that represent various characteristics of traffic flow are extracted. Second, to identify the most important input features for forecasting modeling, a data-driven feature selection strategy is used. Third, a combined forecasting model based on the BCRF algorithm is trained with the selected features. An empirical study on six data sets collected from three different categories of urban roads (expressways, major arterials, and minor arterials) shows that the proposed approach achieves the lowest average MAE and RMSE on six stations, while it achieves the second-best average performance in relation to MAPE. Its ability to identify informative features by feature extraction and feature selection gives the proposed approach excellent forecasting performance and interpretability.

TABLE 5 Computation Time of Competing Models

Model	Time	Ex-1	Ex-2	MaA-1	MaA-2	MiA-1	MiA-2
NAÏVE	TT	0.03	0.05	0.05	0.05	0.09	0.10
	FT	0.01	0.01	0.01	0.01	0.01	0.01
ARIMA	TT	0.09	1.18	1.87	0.13	0.12	0.85
	FT	0.07	0.06	0.07	0.06	0.07	0.06
NNR_{uv}	TT	NA	NA	NA	NA	NA	NA
	FT	0.18	0.19	0.33	0.29	0.32	0.35
$CART_{uv}$	TT	0.2	0.22	0.27	0.28	0.26	0.26
	FT	0.01	0.01	0.01	0.01	0.01	0.01
$BCCART_{uv}$	TT	0.58	0.66	0.74	0.80	0.70	0.71
	FT	0.01	0.01	0.01	0.01	0.01	0.01
RF_{uv}	TT	109.74	110.30	114.19	109.81	109.64	109.30
	FT	0.09	0.08	1.28	0.10	0.08	0.07
NNR_{mv}	TT	NA	NA	NA	NA	NA	NA
	FT	0.53	0.67	1.61	1.46	2.75	1.84
$CART_{mv}$	TT	0.98	1.09	1.45	1.44	1.41	1.43
	FT	0.01	0.01	0.01	0.01	0.01	0.01
$BCCART_{mv}$	TT	2.81	3.18	4.00	4.11	3.74	3.88
	FT	0.01	0.02	0.01	0.01	0.02	0.01
RF_{mv}	TT	410.53	424.41	443.56	413.55	436.08	406.79
	FT	0.14	0.17	0.14	0.16	0.14	0.13
$FSRF_{mv}$	TT	173.09	144.53	169.71	175.59	165.53	188.67
	FT	0.09	0.12	0.11	0.10	0.19	0.11
$FSBCRF_{mv}$	TT	367.06	304.24	342.37	361.30	356.51	368.56
	FT	0.29	0.36	0.31	0.32	0.28	0.29

NOTE: Time in seconds. NA = not available; TT = training time; FT = forecasting time.

The feature selection results in this study suggest that simply choosing the first several lagged variables as input features may not be a wise decision in traffic forecasting modeling. To achieve satisfactory forecast accuracy, the input features and the algorithm should be carefully treated and designed. The forecasting framework proposed in this paper is useful in a variety of applications, so different feature selection strategies and diverse forecasting models can be easily integrated into the framework without the need for major alterations. Moreover, although only traffic volume forecasting is considered in this paper, the proposed approach can also be applied to forecast other available traffic variables, such as speed and travel time.

The main disadvantage of the proposed approach lies in its relatively poor efficiency during the training process. However, given that offline training of the forecasting model is perfectly feasible, this deficiency should not be a significant concern for practical applications. Moreover, since RF is a type of parallel learning algorithm, the training cost will be reduced further if a distributed computing scheme is adopted. Further studies could explore the use of additional competing methods to verify the effectiveness of the proposed model, such as seasonal ARIMA, support vector regression, and neural networks. Adding more features that can be mined and poured into the candidate feature pool, such as weather type, could also strengthen the predictive power of the proposed forecasting approach. In the future, the authors would also like to test the feasibility of the approach on other road types, such as freeways and rural highways.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support received from the China Xinjiang Autonomous Region Science and Technology Support Project. The authors also thank the Kunshan Traffic Management Bureau and the Jiangsu Intelligent Transportation Systems Co., Ltd, for providing the traffic flow data for this study.

REFERENCES

1. Vlahogianni, E.I., J.C. Golias, and M.G. Karlaftis. Short-Term Traffic Forecasting: Overview of Objectives and Methods. *Transport Reviews,* Vol. 24, No. 5, 2004, pp. 533–557. https://doi.org/10.1080/0144164042000195072.
2. Wu, Y.J., F. Chen, C.T. Lu, and S. Yang. Urban Traffic Flow Prediction Using a Spatio-Temporal Random Effects Model. *Journal of Intelligent Transportation Systems,* Vol. 20, No. 3, pp. 1–12. https://doi.org/10.1080/15472450.2015.1072050.
3. Ahmed, M.S., and A.R. Cook. Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques. *Transportation Research Record,* No. 722, 1979, pp. 1–9.
4. Davis, G.A., N.L. Nihan, M.M. Hamed, and L.N. Jacobson. Adaptive Forecasting of Freeway Traffic Congestion. *Transportation Research Record,* No. 1287, 1990, pp. 29–33.
5. Hamed, M.M., H.R. Al-Masaeid, and Z.M. Bani Said. Short-Term Prediction of Traffic Volume in Urban Arterials. *Journal of Transportation Engineering,* Vol. 121, No. 3, 1995, pp. 249–254. https://doi.org/10.1061/(ASCE)0733-947X(1995)121:3(249).
6. Williams, B.M., P.K. Durvasula, and D.E. Brown. Urban Freeway Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models. *Transportation Research Record,* No. 1644, 1998, pp. 132–141. https://doi.org/10.3141/1644-14.
7. 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. https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(664).
8. Kamarianakis, Y., and P. Prastacos. Forecasting Traffic Flow Conditions in an Urban Network: Comparison of Multivariate and Univariate Approaches. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1857, 2003, pp. 74–84. http://dx.doi.org/10.3141/1857-09.
9. Xia, J.X., M. Chen, and W. Huang. A Multistep Corridor Travel-Time Prediction Method Using Presence-Type Vehicle Detector Data. *Journal of Intelligent Transport Systems,* Vol. 15, No. 2, 2011, pp. 104–113. https://doi.org/10.1080/15472450.2011.570114.
10. Stephanedes, Y.J., P.G. Michalopoulos, and R.A. Plum. Improved Estimation of Traffic Flow for Real-Time Control. *Transportation Research Record,* No. 795, 1981, pp. 28–39.
11. Okutani, I., and Y.J. Stephanedes. Dynamic Prediction of Traffic Volume Through Kalman Filtering Theory. *Transportation Research Part B: Methodological,* Vol. 18, No. 1, 1984, pp. 1–11. https://doi.org/10.1016/0191-2615(84)90002-X.
12. Stathopoulos, A., and M.G. Karlaftis. A Multivariate State Space Approach for Urban Traffic Flow Modeling and Prediction. *Transportation Research Part C: Emerging Technologies,* Vol. 11, No. 2, 2003, pp. 121–135. https://doi.org/10.1016/S0968-090X(03)00004-4.
13. Davis, G.A., and N.L. Nihan. Nonparametric Regression and Short-Term Freeway Traffic Forecasting. *Journal of Transportation Engineering,* Vol. 117, No. 2, 1991, pp. 178–188. https://doi.org/10.1061/(ASCE)0733-947X(1991)117:2(178).
14. Smith, B.L., B.M. Williams, and R.K. Oswald. Comparison of Parametric and Nonparametric Models for Traffic Flow Forecasting. *Transportation Research Part C: Emerging Technologies,* Vol. 10, No. 4, 2002, pp. 303–321. https://doi.org/10.1016/S0968-090X(02)00009-8.
15. Clark, S. Traffic Prediction Using Multivariate Nonparametric Regression. *Journal of Transportation Engineering,* Vol. 129, No. 2, 2003, pp. 161–168. https://doi.org/10.1061/(ASCE)0733-947X(2003)129:2(161).
16. Zheng, Z.D., and D.C. Su. Short-Term Traffic Volume Forecasting: A *k*-Nearest Neighbor Approach Enhanced by Constrained Linearly Sewing Principle Component Algorithm. *Transportation Research Part C: Emerging Technologies,* Vol. 43, 2014, pp. 143–157. https://doi.org/10.1016/j.trc.2014.02.009.
17. Dell'Acqua, P., F. Bellotti, R. Berta, and A. De Gloria. Time-Aware Multivariate Nearest Neighbor Regression Methods for Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 16, No. 6, 2015, pp. 3393–3402. https://doi.org/10.1109/TITS.2015.2453116.
18. 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. https://doi.org/10.1016/j.eswa.2008.07.069.
19. Wei, D.L., and H.C. Liu. An Adaptive Margin Support Vector Regression for Short Term Traffic Flow Forecast. *Journal of Intelligent Transport Systems,* Vol. 17, No. 4, 2013, pp. 317–327. https://doi.org/10.1080/15472450.2013.771107.
20. Jeong, Y.S., Y.J. Byon, M.M. Castro-Neto, and S.M. Easa. Supervised Weighting-Online Learning Algorithm for Short-Term Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 14, No. 4, 2013, pp. 1700–1707. https://doi.org/10.1109/TITS.2013.2267735.
21. Stathopoulos, A., and M.G. Karlaftis. Spectral and Cross-Spectral Analysis of Urban Traffic Flows. In *Proceedings of Intelligent Transportation Systems Conference,* 2001, pp. 820–825.
22. Vlahogianni, E.I., M.G. Karlaftis, and J.C. Golias. Optimized and Meta-Optimized Neural Networks for Short-Term Traffic Flow Prediction: A Genetic Approach. *Transportation Research Part C: Emerging Technologies,* Vol. 13, No. 3, 2005, pp. 211–234. https://doi.org/10.1016/j.trc.2005.04.007.
23. Hamad, K., M.T. Shourijeh, E. Lee, and A. Faghri. Near-Term Travel Speed Prediction Utilizing Hilbert–Huang Transform. *Journal of Computer-Aided Civil and Infrastructure Engineering,* Vol. 24, No. 8, 2009, pp. 551–576. https://doi.org/10.1111/j.1467-8667.2009.00620.x.
24. Dunne, S., and B. Ghosh. Regime-Based Short-Term Multivariate Traffic Condition Forecasting Algorithm. *Journal of Transportation Engineering,* Vol. 138, No. 4, 2012, pp. 455–466. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000337.
25. Kumar, K., M. Parida, and V.K. Katiyar. Short Term Traffic Flow Prediction in Heterogeneous Condition Using Artificial Neural Network. *Transport,* Vol. 30, No. 4, 2013, pp. 1–9.
26. Zhu, J.Z., J.X. Cao, and Y. Zhu. Traffic Volume Forecasting Based on Radial Basis Function Neural Network with the Consideration of Traffic Flows at the Adjacent Intersections. *Transportation Research Part C: Emerging Technologies,* Vol. 47, No. 2, 2014, pp. 139–154. https://doi.org/10.1016/j.trc.2014.06.011.

27. Head, K.L. Event-Based Short-Term Traffic Flow Prediction Model. *Transportation Research Record,* No. 1510, 1995, pp. 45–52.
28. Lippi, M., M. Bertini, and P. Frasconi. Short-Term Traffic Flow Forecasting: An Experimental Comparison of Time-Series Analysis and Supervised Learning. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 14, No. 2, 2013, pp. 871–882. https://doi.org/10.1109/TITS.2013.2247040.
29. Ma, T., Z. Zhou, and B. Abdulhai. Nonlinear Multivariate Time-Space Threshold Vector Error Correction Model for Short Term Traffic State Prediction. *Transportation Research Part B: Methodological,* Vol. 76, 2015, pp. 27–47. https://doi.org/10.1016/j.trb.2015.02.008.
30. Caruana, R., N. Karampatziakis, and A. Yessenalina. An Empirical Evaluation of Supervised Learning in High Dimensions. In *Proceedings of the 25th International Conference on Machine Learning,* Helsinki, Finland, 2008, pp. 96–103. https://doi.org/10.1145/1390156.1390169.
31. Vlahogianni, E.I., M.G. Karlaftis, and J.C. Golias. Short-Term Traffic Forecasting: Where We Are and Where We're Going. *Transportation Research Part C: Emerging Technologies,* Vol. 43, No. 1, 2014, pp. 3–19. https://doi.org/10.1016/j.trc.2014.01.005.
32. Zhang, Y.L., and Z.R. Ye. Short-Term Traffic Flow Forecasting Using Fuzzy Logic System Methods. *Journal of Intelligent Transport Systems,* Vol. 12, No. 3, 2008, pp. 102–112. https://doi.org/10.1080/15472450802262281.
33. Tan, M.-C., S.C. Wong, J.-M. Xu, Z.-R. Guan, and P. Zhang. An Aggregation Approach to Short-Term Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 10, No. 1, 2009, pp. 60–69. https://doi.org/10.1109/TITS.2008.2011693.
34. Wang, J., W. Deng, and Y.T. Guo. New Bayesian Combination Method for Short-Term Traffic Flow Forecasting. *Transportation Research Part C: Emerging Technologies,* Vol. 43, 2014, pp. 79–94. https://doi.org/10.1016/j.trc.2014.02.005.
35. Guyon, I., and A. Elisseeff. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research,* Vol. 3, 2003, pp. 1157–1182.
36. Chandrashekar, G., and F. Sahin. A Survey on Feature Selection Methods. *Computers and Electrical Engineering,* Vol. 40, No. 1, 2014, pp. 16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024.
37. Genuer, R., J.M. Poggi, and C. Tuleau-Malot. Variable Selection Using Random Forests. *Pattern Recognition Letters,* Vol. 31, No. 14, 2010, pp. 2225–2236. https://doi.org/10.1016/j.patrec.2010.03.014.
38. Breiman, L. Random Forests. *Machine Learning,* Vol. 45, No. 1, 2001, pp. 5–32. https://doi.org/10.1023/A:1010933404324.
39. Breiman, L.I., J.H. Friedman, R.A. Olshen, and C.J. Stone. Classification and Regression Trees (CART). *Biometrics,* Vol. 40, No. 3, 1984, pp. 17–23.
40. Nguyen, T.-T., J.Z. Huang, and T.T. Nguyen. Two-Level Quantile Regression Forests for Bias Correction in Range Prediction. *Machine Learning,* Vol. 101, No. 1–3, 2015, pp. 325–343. https://doi.org/10.1007/s10994-014-5452-1.
41. Zhang, G., and Y. Lu. Bias-Corrected Random Forests in Regression. *Journal of Applied Statistics,* Vol. 39, No. 1, 2012, pp. 151–160. https://doi.org/10.1080/02664763.2011.578621.
42. Breiman, L. Bagging Predictors. *Machine Learning,* Vol. 24, No. 2, 1996, pp. 123–140. https://doi.org/10.1007/BF00058655.
43. Xia, J., Q. Nie, W. Huang, and Z. Qian. Reliable Short-Term Traffic Flow Forecasting for Urban Roads: Multivariate Generalized Autoregressive Conditional Heteroscedasticity Approach. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2343, 2013, pp. 77–85. https://doi.org/10.3141/2343-10.
44. Washington, S., M.G. Karlaftis, and F.L. Mannering. *Statistical and Econometric Methods for Transportation Data Analysis.* Chapman and Hall/CRC, Boca Raton, Fla., 2003. https://doi.org/10.1201/9780203497111.
45. Guo, J.H., J.X. Xia, and B.L. Smith. Kalman Filter Approach to Speed Estimation Using Single Loop Detector Measurements Under Congested Conditions. *Journal of Transportation Engineering,* Vol. 135, No. 12, 2009, pp. 927–934. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000071.
46. Hyndman, R. *Forecasting Functions for Time Series and Linear Models.* R Package Version 3.2.5, 2016. http://github.com/robjhyndman/forecast.
47. Li, S.Q. *Fast Nearest Neighbor Search Algorithms and Applications.* R Package Version 3.2.5, 2013. https://cran.r-project.org/web/packages/FNN/index.html.
48. Therneau, T., B. Atkinson, and B. Ripley. *Recursive Partitioning and Regression Trees.* R Package Version 3.2.5, 2015. https://cran.r-project.org/web/packages/rpart/.
49. Liaw, A. *Breiman and Cutler's Random Forests for Classification and Regression.* R Package Version 3.2.5, 2015.

The Standing Committee on Artificial Intelligence and Advanced Computing Applications peer-reviewed this paper.