IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 21, 2021, accepted June 3, 2021, date of publication June 8, 2021, date of current version June 22, 2021.

*Digital Object Identifier 10.1109/ACCESS.2021.3087658*

# Expressway Exit Station Short-Term Traffic Flow Prediction With Split Traffic Flows According Originating Entry Stations

**HOUSONG RUAN[1], BANGYU WU[1], (Member, IEEE), BIN LI[2], ZHU CHEN[1], AND WEIGUO YUN[3]**

[1]School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China
[2]School of Automobile, Chang'an University, Xi'an 710046, China
[3]Zhejiang Geely Holding Group Company Ltd., Hangzhou 310051, China

Corresponding author: Bin Li (binli@chd.edu.cn)

This work was supported in part by the Natural Science Basic Research Program of Shaanxi under Program 2020JM-018, in part by the National Research Program for Key Issues in Air Pollution Control of China under Grant DQGG0207-03, and in part by the National Key Research Project of China under Grant 2017YFC0212103.

**ABSTRACT** As an essential component of Intelligent Transportation Systems (ITS), short-term traffic flow prediction is a key step to anticipate traffic congestion. Due to the availability of massive traffic data, data-driven methods with a variety of features have been applied widely to improve the traffic flow prediction. China has the longest total length of expressways in the world and there is significant information recorded when vehicles enter and exit the expressway. In this paper, we collect data at an expressway exit station in Shanghai, split the data according to its originating entry stations and predict the corresponding exit station traffic flow using the multi split traffic flows. First, the original records are collected, preprocessed, split, aggregated and normalized. Second, the Long Short-Term Memory (LSTM) model is applied to learn from the features of the overall flow and split traffic flows to predict the overall exit flow. The baselines are models which only overall flow information is considered. Compared with the baselines, in other models, the split flows according entry stations are also considered for prediction. Finally, the LSTM model is made comparison with the Convolutional LSTM(ConvLSTM), the K-Nearest Neighbor (KNN) model and the Support Vector Regression (SVR) model. When the information of overall flow and 6 split traffic flows is used and *step* is set to 11 (with 5 minute aggregation), the model prediction performs best. Compared with the best result of LSTM baseline model, the improvement of prediction accuracy is up to 5.48 percent by Mean Absolute Error (MAE).

**INDEX TERMS** Traffic flow prediction, LSTM model, feature selection, split flows.

## I. INTRODUCTION

China is experiencing greater urbanization and traffic congestion is one of the major challenges [1]. Serious traffic congestion and the resulting adverse influence on traffic safety and environmental conditions have received considerable attention [2]. To address the problems, Intelligent Transportation Systems (ITS) have been extensively studied for decades and has emerged as an effective way to improve efficiency of transportation [3]. Short-term traffic flow prediction has become one of the major research fields in ITS [4]. Accurate real-time short term traffic flow predictions help traffic participants choose appropriate travel routes and provide an effective control strategy for traffic managers to relieve traffic congestion [5]. It also reduces pollution and improves the safety of transportation.

With the rapid development of China's economy, the mileage of expressways and the number of vehicles have increased tremendously [6], [7]. Besides the number of vehicles, there is copious data which could be collected from expressway stations [8], like vehicle type [9], entry time, entry station, exit time and exit station. However, most traffic flow prediction models mainly consider the overall flow information alone. Some studies have analyzed the impacts of other information, such as upstream and downstream traffic flow [10], [11], traffic flow at related sites [12]–[14], traffic

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wang.

flow at the link-level [1] and traffic flow of various vehicle types [10]. Few studies have tried to predict the overall flow with the split flow information. By split traffic flow, we mean the proportion of traffic exiting a specific exit station that originates from a given entry station. The overall flow is composed by the split flows with different features. Vehicles from various entry stations arrive at an exit site with different travel times and quantity. In this matter, the split flows of different entry stations contribute various patterns to the overall flow. It's believed that the accuracy of traffic flow predictions can be improved, if incorporating the split flows as inputs. Moreover, if the most effective split flows rather than all flows are used as the input for a model, it will achieve optimal predictive capability.

At present, two kinds of techniques have been applied in traffic flow prediction, one is the parametric model and the other is the nonparametric model [15]. Parametric models refer to models with fixed structures based on some assumptions [16]. Nonparametric models refer to models without fixed structures, such as K-Nearest Neighbor (KNN), Artificial Neural Network (ANN) and so on [16].

Parametric models contain Autoregressive Integrated Moving Average (ARIMA) model [17], Kalman Filtering [18] and so on. ARIMA and its variants are the most commonly used parametric models in traffic flow prediction [16], like Autoregressive Integrated Moving Average with Generalized Autoregressive Conditional Heteroscedasticity (ARIMA-GARCH) [19] and Seasonal ARIMA (SARIMA) [20]. Parametric models are more efficient than nonparametric models [15].

In recent years, nonparametric models have received intensive attention in traffic flow prediction. The change in road traffic flow is a real-time and ever-changing process [21], and nonparametric models are suited for nonlinear and stochastic traffic data [4], [15]. Using large data sets, the nonparametric models can potentially improve the accuracy of predictions [1]. Lv [22] applied a deep architecture model to use autoencoders as building blocks to represent traffic flow features for prediction. It had better performance than Back Propagation Network (BP), Random Walk (RW), Support Vector Machine (SVM) and Radial Basis Function Network (RBFN). Zhu [23] used Radial Basis Function Neural Networks (RBFNN) to predict the urban traffic flow. And the traffic volumes of adjacent intersections are used as inputs to obtain more accurate forecasting result. Zeng [24] proposed a hybrid model, which estimates the weights of the Multilayer Artificial Neural Network (MLANN) and the parameters of ARIMA model. The MLANN model is used to analyze the nonlinear part of traffic flow time series and the ARIMA model is developed to model the residuals from the ANN model. Wang [25] proposed a hybrid model of SVM and KNN for short-term freeway exiting volume prediction. Compared with a single KNN method or Support Vector Regression (SVR) method, the hybrid model improves the stability of prediction results.

As a nonparametric model, ANN plays a crucial role in traffic volume forecast. However, the traffic flow is sequential data in nature. Multilayer Perceptron (MLP) is the most widely used form of ANN, but its output depends only on the current input and not on any past inputs [26]. To relax this condition, the Recurrent Neural Network (RNN) was applied to forecast the traffic flow. As a special kind of ANN, there are self-connected hidden layers on RNN so that the RNN can map information of previous inputs to each output [27]. But when a RNN model is trained to predict the traffic flow and the Back Propagation Through Time (BPTT) is used to update weights, the gradients of the RNN may vanish or explode with time lags increasing [16]. As a particular type of RNN, the Long Short-Term Memory (LSTM) model was proposed to overcome the disadvantages [10]. Shao [28] explored the applications of LSTM model in short-term traffic prediction. Compared with RW, SVR, Wavelet Neural Network (WNN) and Stacked AutoEncoder (SAE), LSTM model has the smallest error in both Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). Tian [29] applied the LSTM to predict the traffic flow with 15-min, 30-min, 45-min and 60-min intervals, and the prediction results showed that the LSTM demonstrated more excellent generalization capability than RW, SVM, SAE and single hidden layer Feed Forward Neural Network (FFNN). Wang [30] made a traffic flow prediction by using the sampled GPS data and the results showed that the LSTM and Gated Recursive Unit (GRU) performed better than ARIMA and SVR. The LSTM model has more accurate prediction results than GRU. Luo [13] used KNN to choose related neighboring sites. Then the traffic flow data of these sites were used as the inputs of LSTM. Wei [14] used the AutoEncoder to obtain the internal relationship of traffic flow and applied the LSTM network for traffic flow prediction. Some study use other data information about the traffic flow as the input of the LSTM model, like upstream and downstream traffic flows [14], weather conditions [1], [31] and traffic flows at related sites [13]. Zheng [1] proposed a deep and embedding learning approach (DELA) that consisted of an embedding component, a Convolutional Neural Network (CNN) component and a LSTM component. The impacts of traffic flow, route structures, weather conditions and special events were considered together. Kang [10] analyzed the effects of various inputs on the LSTM prediction performances. The inputs of each model are various combinations of data sets that contain the upstream and downstream traffic flows, speed and occupancy.

Feature/Variable selection contributes to the improvement of traffic flow prediction. It contains two methods: filter methods and wrapped methods. The filter methods select features/variables by ranking them and filtering out the least promising features/variables [32]. The wrapped methods analyze the prediction performance of a given model to select a subset of features/variables [32]. Different combinations of features were applied to select the most effective features, like [10] and [33]. Wu [33] proposed a novel

short-term traffic flow prediction approach based on the combination of CNN and LSTM (CLTFP). The spatial features, short-term temporal features and periodic features were generated from well-trained model CLTFP to train several Lasso models. And the features were combined variously to conduct a comparison among prediction results of Lasso models. The Lasso model with 3 features outperforms other Lasso models. Yan [34] used Pearson correlation coefficient to select the flows at neighboring sites. Then the spatial and temporal features of flows were extracted to predict the traffic flow. Zhang [35] proposed a Spatio-temporal feature selection algorithm (STFSA). The ANN model and CNN model were used with the algorithm. For a model using STFSA, it achieved better performances but spent more than ten times of the training time of the non-STFSA model.

To fill the gap of the previous studies, we apply the LSTM model to learn from the history feature of the overall flow and split flows according entry stations to forecast the exit flow at the exit station of expressway. The baselines are the models in which only the overall flow information is considered. In this paper, the overall flow is split according the entry stations and ranked according the traffic of vehicles. The spilt flows with little cross-correlation are not considers as the input of models. Rest split flows are used as the input of model. The main contributions of this study are as follows:

(I) Consider the impact of the split flows and make full use of different traffic flow features.

(II) Analyze the effects of various inputs on the overall flow predictions compared with the baselines. So that we can ensure the split flows could improve the model.

We note that a conference version has been presented [36]. Compared to previous work, we give detailed analysis on the feature selection, data preprocessing and model comparison in this paper. First, the trend of traffic flow is described accurately. Second, the feature selection is explained by cross-correlation analysis. Third, process about determining the values of hyperparameters is described. Fourth, the comparison of various models shows that the information of split flows makes models achieve better performance.

The rest of this paper is organized as follows. Section 2 discusses the theory of the methods. Section 3 shows the experiments of predictions. The conclusion is provided in Section 4.

## II. METHODOLOGY

### A. DATA PREPARATION

The overall flow $Y$ could be denoted as $\{Y_1, Y_2, \cdots, Y_T\}$, and $Y_t(t = 1, 2, \ldots, T)$ is the value of the overall flow at timestamp $t$. The split flows are expressed as $\{X_1, X_2, \cdots, X_T\}$ where $X_t = (x_{1t}, x_{2t}, \cdots, x_{Nt})$. $x_{it}(i = 1, 2, \ldots, N)$ is the value of split traffic flow of entry station $i$ at timestamp $t$. $x_i$ is the split flow of entry station $i$ and denoted as $(x_{i1}, x_{i2}, \cdots, x_{iT})^T$ $i = 1, 2, \cdots, N$.

### B. DATA NORMALIZATION

Before training the LSTM model, the data is normalized into range [0,1] by Min-Max normalization methods. The normalization method is described as eq (1) and eq (2).

$$Y_t^{norm} = \frac{Y_t - Y^{min}}{Y^{max} - Y^{min}} \tag{1}$$

$$x_{it}^{norm} = \frac{x_{it} - x_i^{min}}{x_i^{max} - x_i^{min}} \tag{2}$$

The $Y^{max}$ is the maximum value in the overall flow, and the $Y^{min}$ is the minimum value. The $x_i^{max}$ is the maximum value in the split flow of entry station $i$, and the $x_i^{min}$ is the minimum value. The normalized overall flow $Y^{norm}$ is expressed as $\{Y_1^{norm}, Y_2^{norm}, \cdots, Y_T^{norm}\}$. The normalized split flows $X^{norm}$ are denoted as $\{X_1^{norm}, X_2^{norm}, \cdots, X_T^{norm}\}$ where $X_t^{norm}$ is denoted as $(x_{1t}^{norm}, x_{2t}^{norm}, \cdots, x_{Nt}^{norm})$.

### C. LSTM MODEL

The LSTM layer is composed of input gate, forget gate, output gate, and self-connected memory cell. Input gate is used to enter current data information. Forget gate is used to determine how much previous information should be abandoned. Output gate is used to output data information to the next layer or LSTM layer at next timestamp. Memory cell restores the previous data information. Input gate, forget gate and output gate are activated by the sigmoid function $\sigma(\cdot)$ and three gates' outputs are between 0 and 1. $\sigma(\cdot)$ is defined in eq (3).

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

*Baselines:* When the number of split flows $N$ is 0, it indicates that only the overall flow information is considered in the models. These models are recognized as the baselines. The LSTM model in baseline has been shown in Figure 1. $H$ is the length of hidden state vector in the LSTM layer.

At timestamp $j$, the overall flow information $Y_j^{norm}$ and the hidden state $h_{j-1}$ are mapped to the LSTM layer. The $f_j$ decides how much previous memory should be kept at timestamp $j$. The $i_j$ controls the amount of new information content being added to the current memory [37]. They are calculated as:

$$f_j = \sigma(Y_j^{norm} W_{yf} + h_{j-1} W_{hf} + b_f) \tag{4}$$

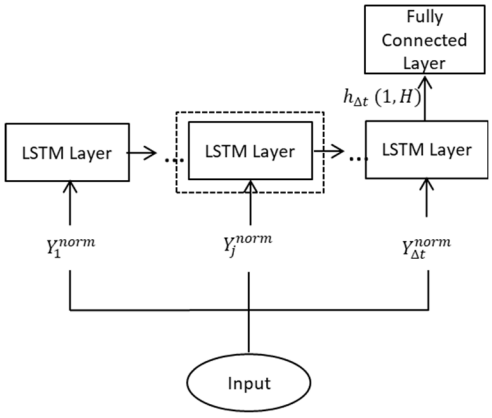
$$i_j = \sigma\left(Y_j^{norm} W_{yi} + h_{j-1} W_{hi} + b_i\right) \tag{5}$$

where $W_{yf} W_{yi} \in R^{1\times H}$, $W_{hf} W_{hi} \in R^{H\times H}$, $b_f b_i \in R^{1\times H}$, and $f_j\ i_j \in R^{1\times H}$. $W$ is the weight matrices between individual units and $b$ is the bias vector.

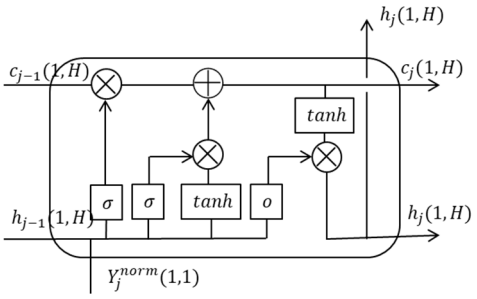
The hidden state $h_j$ is computed by output gate $o_j$ and cell state $c_j$. The $c_j$ and $h_j$ will be mapped to the LSTM layer at timestamp $j + 1$. The cell state is computed as:

$$\tilde{c}_j = tanh\left(Y_j^{norm} W_{yc} + h_{j-1} W_{hc} + b_c\right) \tag{6}$$

$$c_j = f_j \odot c_{j-1} + i_j \odot \tilde{c}_j \tag{7}$$

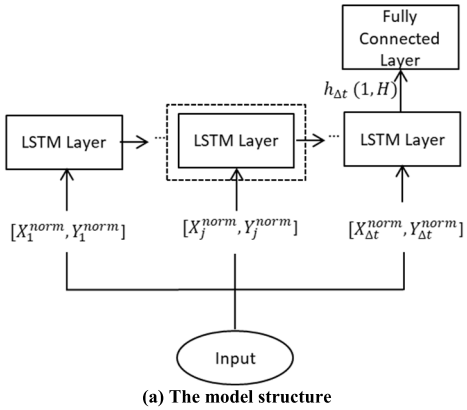



(a) The model structure

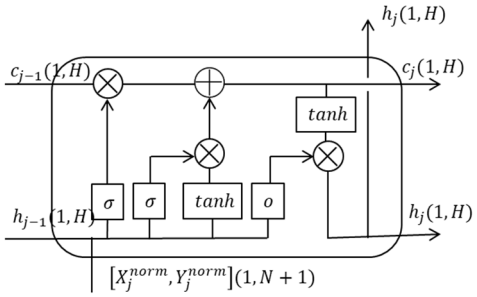



(b) The details of LSTM layer at time $j$

**FIGURE 1.** The LSTM model in baseline.




(a) The model structure




(b) The details of LSTM layer at time $j$

**FIGURE 2.** The LSTM models which consider split flows.

$$o_j = \sigma\left(Y_j^{norm} W_{yo} + h_{j-1} W_{ho} + b_o\right) \quad (8)$$

$$h_j = o_j \odot tanh(c_j) \quad (9)$$

where $W_{yc} W_{yo} \in R^{1\times H}$, $W_{hc} W_{ho} \in R^{H\times H}$, $b_c b_o \in R^{1\times H}$, and $\tilde{c}_j o_j \in R^{1\times H}$.$\odot$ implies the multiplication of corresponding elements of two matrices with the same shape.

When the *step* is set to $\Delta t$, the $h_{\Delta t}$ will be mapped to the fully connected layer. Then, the $\hat{Y}_{\Delta t+1}^{norm}$ is computed as eq (10):

$$\hat{Y}_{\Delta t+1}^{norm} = h_{\Delta t} W_h + b_h \quad (10)$$

where $W_h \in R^{H\times 1}$ and $b_h \in R^{1\times 1}$.

*Models with input split flows:* The models which consider split flows information are compared with the baselines to analyze the effect of the information. The model has been plotted in Figure 2. The formulas are presented as:

$$f_j = \sigma([X_j^{norm}, Y_j^{norm}] W_{[x,y]f} + h_{j-1} W_{hf} + b_f) \quad (11)$$

$$i_j = \sigma\left([X_j^{norm}, Y_j^{norm}] W_{[x,y]i} + h_{j-1} W_{hi} + b_i\right) \quad (12)$$

$$\tilde{c}_j = tanh\left([X_j^{norm}, Y_j^{norm}] W_{[x,y]c} + h_{j-1} W_{hc} + b_c\right) \quad (13)$$

$$c_j = f_j \odot c_{j-1} + i_j \odot \tilde{c}_j \quad (14)$$

$$o_j = \sigma\left([X_j^{norm}, Y_j^{norm}] W_{[x,y]o} + h_{j-1} W_{ho} + b_o\right) \quad (15)$$

$$H_j = o_j \odot tanh(c_j) \quad (16)$$

$$\hat{Y}_{\Delta t+1}^{norm} = h_{\Delta t} W_h + + b_h \quad (17)$$

where $W_{[x,y]f} W_{[x,y]i} W_{[x,y]c} W_{[x,y]o} \in R^{(1+N)\times H}$, $W_{hf} W_{hi} W_{hc} W_{ho} \in R^{H\times H}$, $b_f b_i b_c b_o \in R^{1\times H}$, $f_j i_j \tilde{c}_j o_j \in R^{1\times H}$, $W_h \in R^{H\times 1}$ and $b_h \in R^{1\times 1}$.

In this paper, the object function of model training is set as eq (18).

$$Loss = \frac{1}{m} \sum_{i=1}^{m} |\hat{Y}_i^{norm} - Y_i^{norm}| \quad (18)$$

The Adam optimizer is applied to minimize the object function. Adam optimizer, a modification of stochastic gradient descent (SGD) optimizer with adaptive learning rates, is applied for BPTT [10].

### D. MODEL EVALUATION

The $\hat{Y}_i^{norm}$ is re-scaled into the predicted value $\hat{Y}_i$ and it is computed as follow.

$$\hat{Y}_i = \hat{Y}_i^{norm}\left(Y^{max} - Y^{min}\right) + Y^{min} \quad (19)$$

After the predicted values are re-normalized, we use RMSE and Mean Absolute Error (MAE) to evaluate accuracy of model predictions. They are defined as eq (20) and eq (21):

RMSE

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\hat{Y}_i - Y_i\right)^2} \quad (20)$$

MAE

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{Y}_i - Y_i| \quad (21)$$

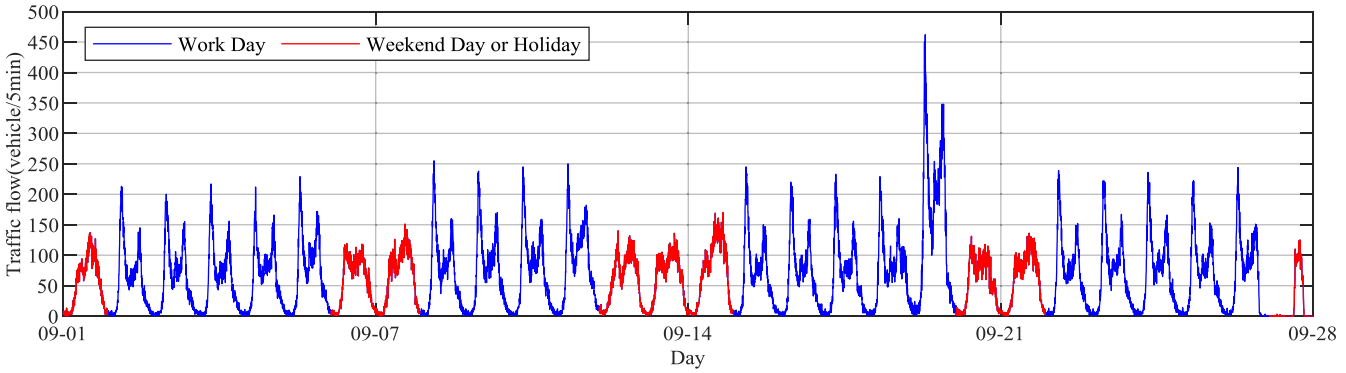



**FIGURE 3.** The overall traffic flow at 5-minute interval.

where $Y_i$ and $\hat{Y}_i$ presents the traffic flow data and predicted traffic flow values in 5-minute interval. The $n$ represents the number of all predict values. The MAE reflects the absolute value of the difference between the predicted value and the actual traffic flow data [1], [38]. The RMSE evaluates the degree of change of data [39]. The lower value of MAE and RMSE imply the more accurate prediction results. In this paper, the MAE value is applied to compare the accuracy of prediction results.

## III. EXPERIMENT

### A. DATA DESCRIPTION AND PREPROCESS

Shanghai is a metropolis in China with developed expressways. The traffic congestion of Shanghai has become serious. ITS has received sustained attention for its ability to solve the issue. As a basic component of ITS, short-term traffic flow prediction is used to help people plan travel and improve traffic management. For this study, we used the expressway Electronic Toll Collection (ETC) traffic flow data from an exit station in Shanghai. The time span of traffic flow data is from 1st September, 2019 to 28th September, 2019.

There are 539,614 records in total and every record includes entry station, exit station, exit time, mileage and vehicle type. There are occasional abnormal records. Abnormal records might have a negative influence on the model prediction and should be filtered out [40]. There are 3 types of abnormal data:

(I) Records with abnormal entry stations. Some codes of entry stations are recorded as 0.

(II) Records with wrong time stamps.

(III) Records with zero mileage.

Excluding abnormal data, the preprocessed data contains 520,502 records. The large number of data points is beneficial to model, so we attempt to obtain more training data as possible [1]. But it will spend more time. In order to balance the efficiency and accuracy of training model, the preprocessed data is aggregated into 5 minutes [1]. It is shown in Figure 3. The blue line and red line are for workday and weekend respectively. Some observations could be summarized as below:

(I) In each work day, the overall flow has at least two peaks: a morning peak and an afternoon peak. The values of morning peaks are always presented with larger values. In a workday, the morning peak reaches values larger than 200 veh/5min, but the afternoon peak reaches values of about 150 veh/5min.

(II) On some work days, the traffic flow peak is at noon.

(III) On a weekend day and a holiday (from 13rd September to 15th September), peak values are less than those on weekdays. Except on 15th September, peaks reach values less than 150 veh/5min.

(IV) In the early morning and late evening, traffic volume is the lowest.

(V) On 20th September, the peak value in the morning exceeds 450veh/5min in the time period.

(VI) On 28th September, traffic flow is close to 0 most of the day (The expressway may have been closed at this time).

The traffic flow distribution varies between weekend days, holiday and week days. The traffic volume on a weekend day is presented in Figure 4 and the corresponding smooth values (rolling mean) are used to show the trend of traffic volume. The traffic volume rises after 4:00. The morning peaks on a weekend day or holiday occur at about 10 o'clock. Then, the traffic volume declines until about 12 o'clock. The afternoon peak values are reached between 16:00 and 18:00. In Figure 4a, the morning peak value is larger. However, in Figure 4b the afternoon peak value is larger. Finally, the traffic volume drops for the rest of the day.

The traffic flow in different hours of a work day is shown in Figure 5. From 0:00 to 4:00, the traffic volume is close to 0. After 4:00, the traffic volume fluctuates dramatically. In a workday, the traffic flow peak comes between 6 o'clock and 8 o'clock on each work day morning. Typically, the peaks in the afternoon are reached between 16:00 and 18:00. The peak values are smaller than those in the morning. Finally, the traffic volume declines in the rest of the day.

According to originating entry stations, the records are split to construct the multi traffic flows. In total, the vehicles are from 102 entry stations, and 102 flows could be constructed. The patterns of various split traffic flows are quite different. Selecting the relevant information is the key point in this study and the Pearson correlation coefficient is recognized

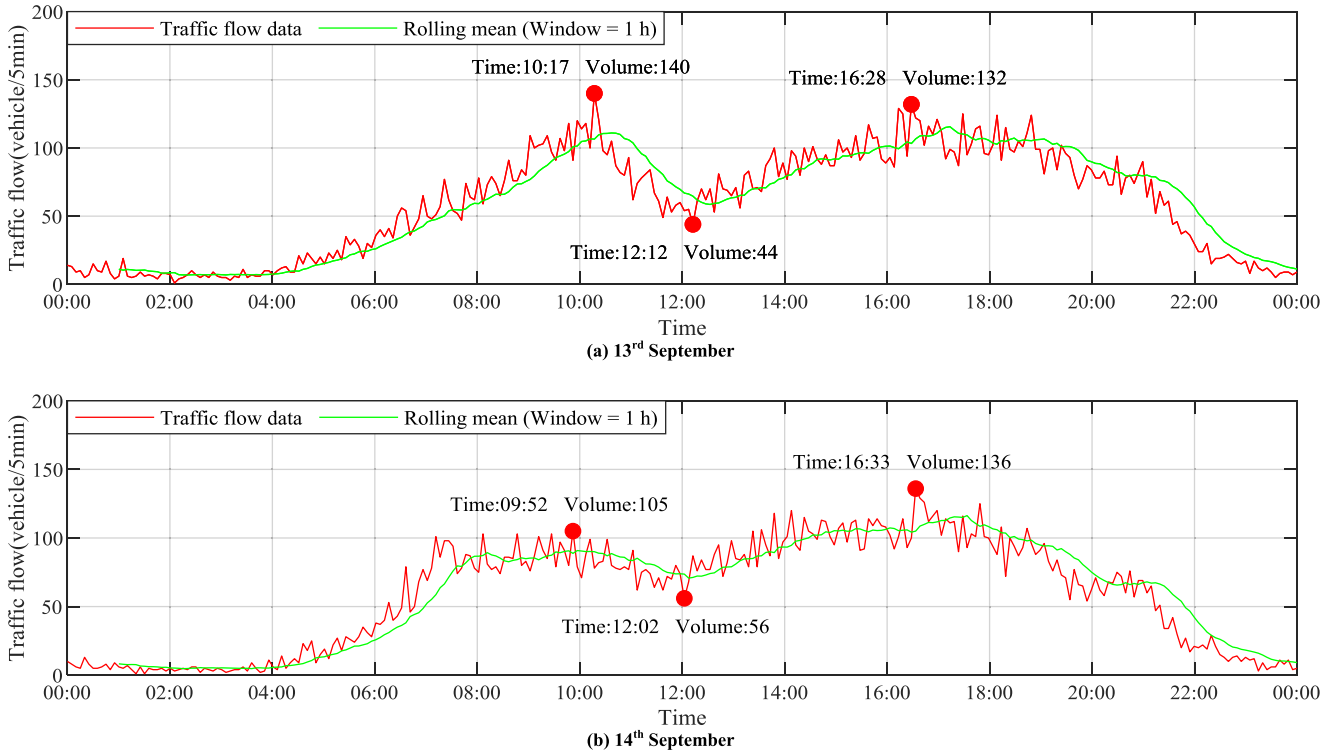



**FIGURE 4.** Typical traffic flow on weekend or holiday.

**TABLE 1.** The overall traffic flow and the split flows.

(a) Number of Vehicles in first 3 weeks

| Station | Number of Vehicles | Rate | Station | Number of Vehicles | Rate |
|---|---|---|---|---|---|
| Exit Station | 341,942 | 1. 0000 | A8 | 10,631 | 0. 0311 |
| A1 | 63,979 | 0. 1871 | A9 | 9,497 | 0. 0278 |
| A2 | 53,906 | 0. 1576 | A10 | 8,985 | 0. 0263 |
| A3 | 42,296 | 0. 1237 | A11 | 7,670 | 0. 0224 |
| A4 | 37,695 | 0. 1102 | A12 | 5,813 | 0. 0170 |
| A5 | 26,386 | 0. 0772 | A13 | 5,478 | 0. 0160 |
| A6 | 16,609 | 0. 0486 | A14 | 5,045 | 0. 0148 |
| A7 | 14,016 | 0. 0410 | -- | -- | -- |

(b) Number of Vehicles from 1st September to 28th September

| Station | Number of Vehicles | Rate | Station | Number of Vehicles | Rate |
|---|---|---|---|---|---|
| Exit Station | 520,502 | 1. 0000 | A8 | 16,335 | 0. 0314 |
| A1 | 98,121 | 0. 1885 | A9 | 14,173 | 0. 0273 |
| A2 | 81,744 | 0. 1570 | A10 | 13,526 | 0. 0260 |
| A3 | 64,626 | 0. 1242 | A11 | 10,911 | 0. 0210 |
| A4 | 57,584 | 0. 1106 | A12 | 8,432 | 0. 0162 |
| A5 | 40,146 | 0. 0771 | A13 | 8,255 | 0. 0159 |
| A6 | 25,640 | 0. 0493 | A14 | 7,655 | 0. 0147 |
| A7 | 21,262 | 0. 0408 | -- | -- | -- |

as a criterion to assess the strength of relevance [35]. The Pearson correlation is obtained by eq (22):

$$\rho(A,B) = \frac{Cov(A,B)}{\sigma_A \sigma_B} = \frac{\sum_{i=1}^{n}(a_i - \bar{a})\sum_{i=1}^{n}(b_i - \bar{b})}{\sqrt{\sum_{i=1}^{n}(a_i - \bar{a})^2}\sqrt{\sum_{i=1}^{n}(b_i - \bar{b})^2}} \tag{22}$$

where $A = \{a_1, a_2, \cdots, a_n\}$ and $B = \{b_1, b_2, \cdots, b_n\}$ represents two sequence data. $a_i$ and $b_i$ are the sequence values at timestamp $i$. $\bar{a}$ and $\bar{b}$ represent the average values of sequence data. $Cov(A, B)$ is the covariance between $A$ and $B$. $\sigma_A$ and $\sigma_B$ express the standard deviations of the sequence data. The closer $|\rho(A, B)|$ is to 1, the stronger the relevance is.

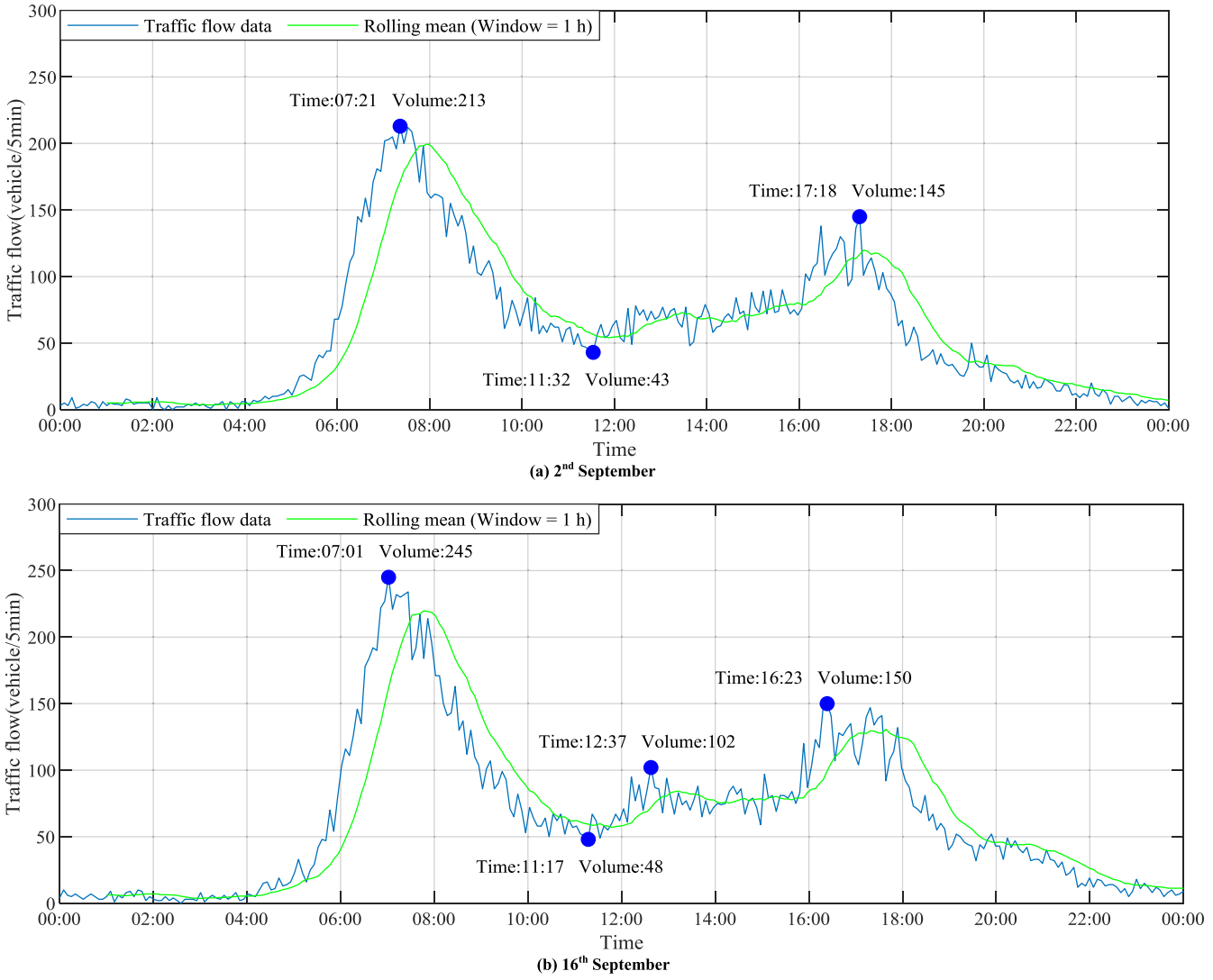


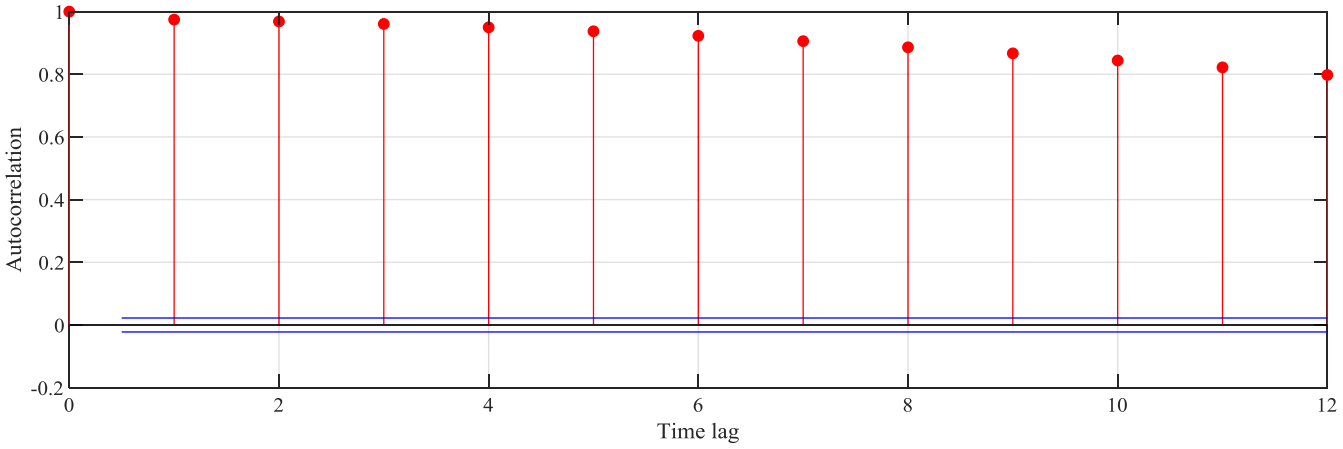

**FIGURE 5.** Typical traffic flow on a work day.

The autocorrelation analysis of the overall flow is presented in Figure 6a and it can be considered that there is a strong temporal correlation in overall traffic flow data. As the time lag increases, the autocorrelation correlation coefficient reduces slowly. When the time lag number is set to 12, the autocorrelation coefficient drops to nearly 0.8. The cross-correlation illustrates the correlation between the overall flow and split flows. When time lag is set to $\Delta t$, the values of cross-correlation is the correlation coefficient between overall flow, from time $\Delta t + 1$ to time $T$ and split flows, from time 1 to time $T - \Delta t$. Figure 6b and Figure 6c show that the flows with larger vehicle numbers are more relevant to the overall flow.

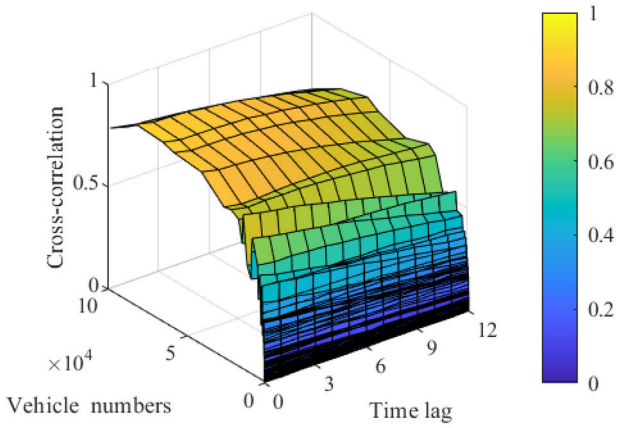
In the first 3 weeks, the total number of vehicles on 14 split flows is approximately 90.08% of the overall flow. From 1st September to 28th September, they are also the largest flows and comprise 89.99% of the overall flow. The details have been provided in Table 1. Their distributions in Shanghai are presented in Figure 7. These split flows are used to train the models and forecast the overall flow.

Two types of flows are presented in Figure 8. As Figure 8a shows, on each work day, the afternoon peak values for A1 are larger than the morning peak values. On contrast, in Figure 8b, the morning peak values for A2 are higher. The pattern of flow from A1 is quite different with the overall flow. So when time lag is 0, the cross-correlation coefficient between overall flow and split flow of A1 is less than that between overall flow and split flow of A2 which is plotted in Figure 8b. The spilt flows provide more details about traffic flow at an exit station. It could be applied in traffic flow prediction.

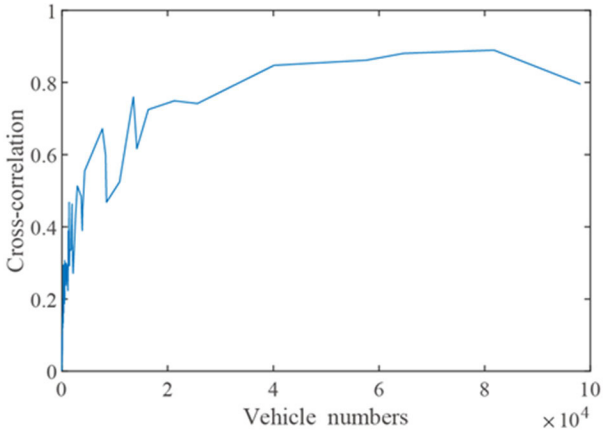
The traffic flow sequence data can be constructed by the follow procedures: First, the original traffic flow data was recorded with detailed information. Second, abnormal records are removed and the preprocessed data is obtained. Third, the preprocessed data is aggregated in 5-min intervals.




(a) Temporal correlation of the overall flow




(b) Cross-correlation




(c) Cross-correlation (Time lag=0)

**FIGURE 6.** Correlation analysis.


**FIGURE 7.** The Distributions of stations in this study [41].

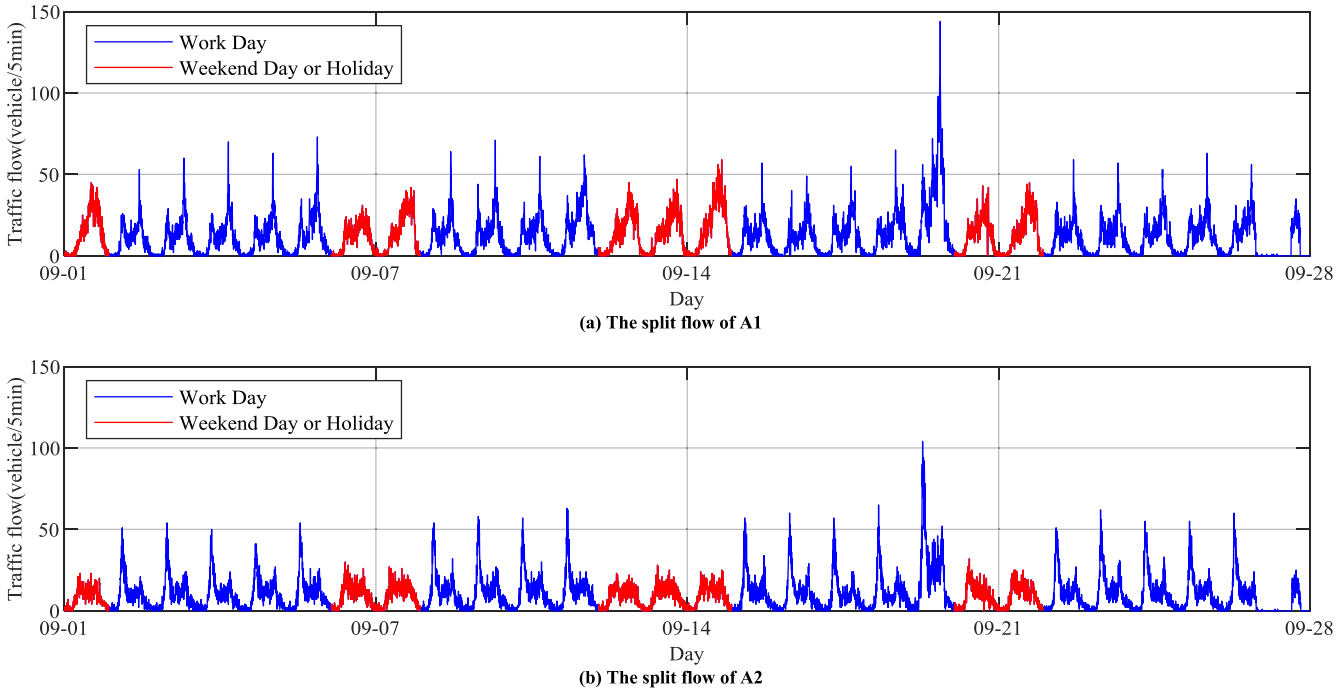



**FIGURE 8.** The examples of split flows.

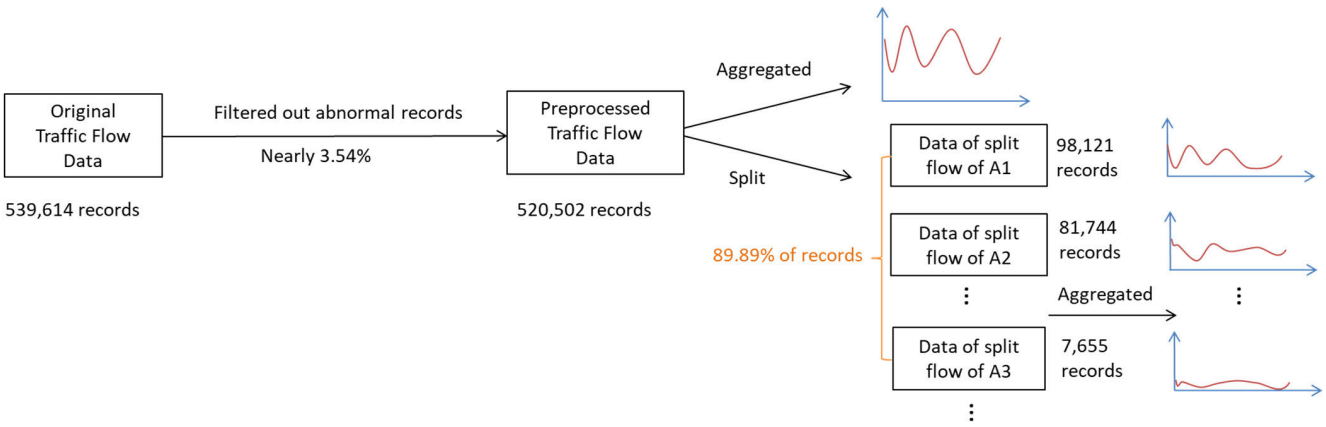


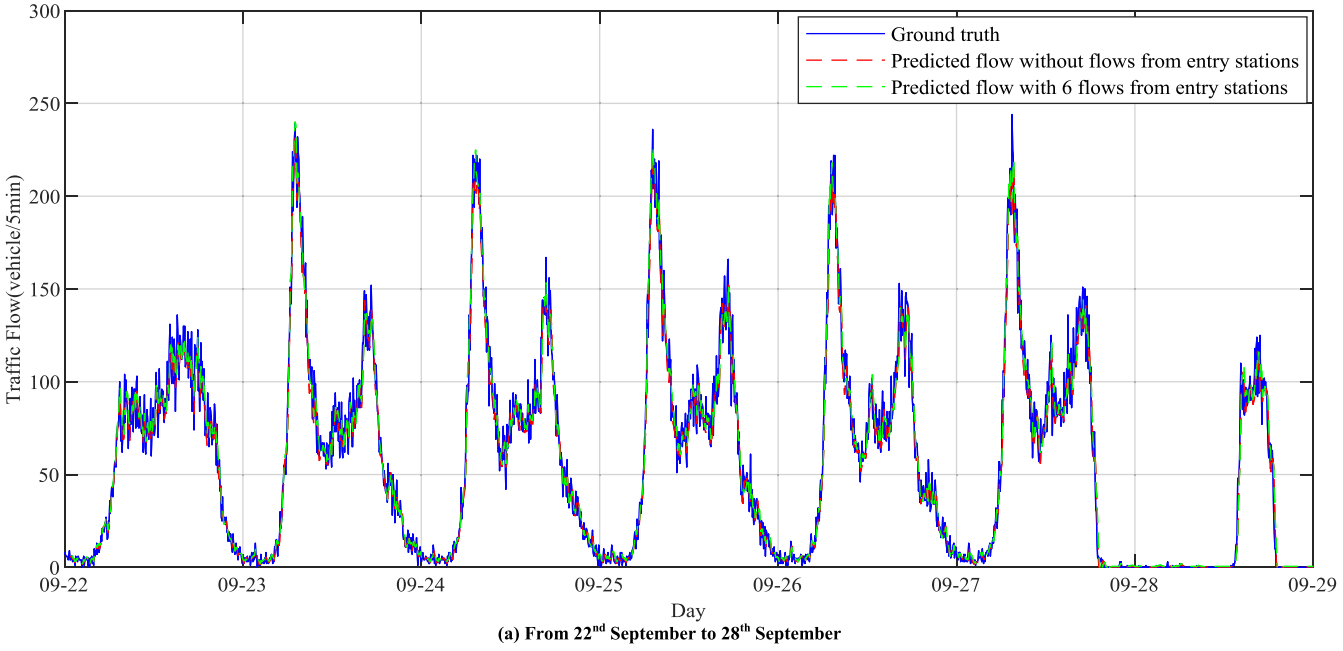

**FIGURE 9.** The process of preprocessing.

Fourth, the preprocessed data is split, and records are summed to obtain the number of vehicles originating from all stations. The stations with largest number of originating vehicles are selected. Finally, the split data is aggregated and split flows are obtained. In this study, the process is summarized in Figure 9.

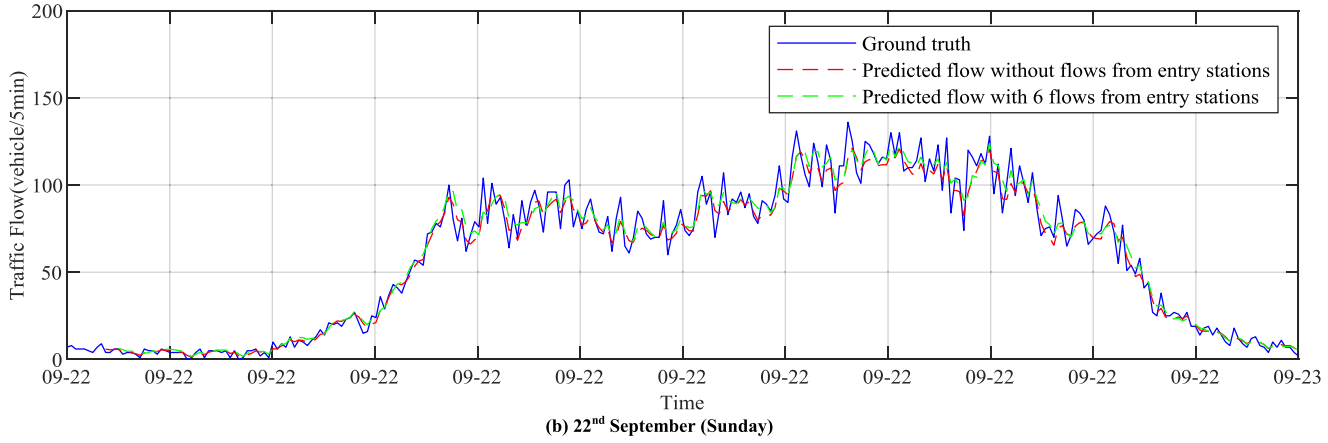
### *B. PREDICTION RESULTS*

The data set is divided into 2 phases: the training phase and the test phase. The training data set contains the traffic flow data from 1$^{st}$ September, 2019 to 21$^{st}$ September, 2019 and the rest data is considered as the test data. The prediction results of models are compared versus the test data set. In this paper, the short-term prediction refers to the prediction of traffic flow in the next 1-hour or less interval. So, *step* is set to the range from 1 to 12, that is, from 5 minutes to 1 hour. The output in the models is the normalized value of overall flow at timestamp $(step+1)$-th. The predicted overall flow values are obtained by reversing the normalization process. In baselines of LSTM models, only the overall flow is considered as input. In other models, the input includes the overall flow and the $N$ split flows. The performances of other models are compared against the baselines.

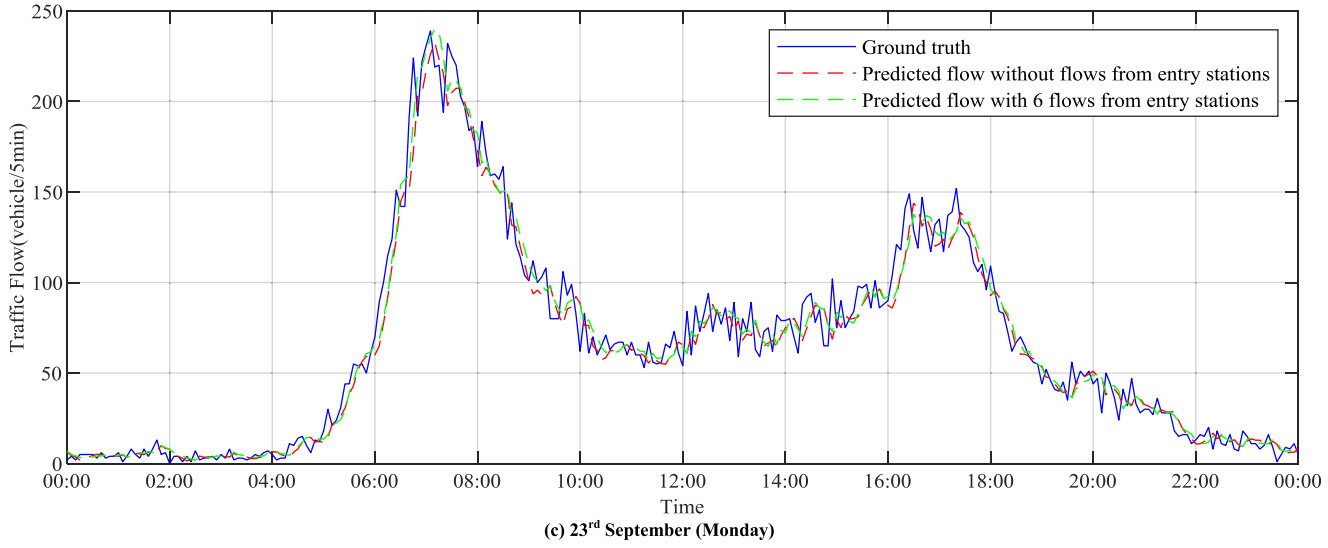
There are some hyperparameters that need to be considered before the LSTM model is trained: the length of hidden vector, the number of epochs and batch_size. When all hyperparameters are fixed, the most complex LSTM model is when the traffic flow number $N$ is set to 14 and *step* to 12. The simplest model is set $N$ to 0 and *step* to 1. The LSTM models are trained using the training data set,




(a) From 22nd September to 28th September




(b) 22nd September (Sunday)




(c) 23rd September (Monday)

**FIGURE 10.** The comparison between traffic flow data and predicted values.

**TABLE 2.** The model performance with various parameters ($N = 0$, $step = 1$).

(a) Hidden vector

| Length | MAE | Length | MAE | Length | MAE |
|---|---|---|---|---|---|
| 2 | 39.9213 | 16 | 23.9471 | 128 | 8.1788 |
| 4 | 38.1886 | 32 | 20.7828 | 256 | 8.8271 |
| 8 | 31.1034 | **64** | **8.1564** | - | - |

(b)Epoch

| Epoch | MAE | Epoch | MAE | Epoch | MAE |
|---|---|---|---|---|---|
| 5 | 8.1333 | 25 | 9.0744 | 45 | 7.9066 |
| 10 | 8.0683 | 30 | 8.4742 | 50 | 7.9275 |
| 15 | 7.9460 | 35 | 7.9037 | - | - |
| **20** | **7.9015** | 40 | 7.9700 | - | - |

(c)Batch_size

| Batch_size | MAE | Batch_size | MAE |
|---|---|---|---|
| 2 | 7.9726 | 32 | 8.6691 |
| 4 | 9.1833 | 64 | 8.0397 |
| 8 | 8.0800 | 128 | 7.9310 |
| **16** | **7.9286** | 256 | 8.0452 |

**TABLE 3.** The model performance with various parameters ($N = 14$, $step = 12$).

(a) Hidden vector

| Length | MAE | Length | MAE | Length | MAE |
|---|---|---|---|---|---|
| 2 | 13.3268 | 16 | 9.0856 | **128** | **7.5760** |
| 4 | 11.4891 | 32 | 8.4015 | 256 | 7.9174 |
| 8 | 7.9180 | 64 | 8.1564 | - | - |

(b) Epoch

| Epoch | MAE | Epoch | MAE | Epoch | MAE |
|---|---|---|---|---|---|
| 5 | 7.6651 | 25 | 6.9808 | 45 | 7.1052 |
| 10 | 7.0171 | 30 | 7.0297 | 50 | 7.0543 |
| 15 | 7.0486 | 35 | 6.9765 | - | - |
| **20** | **6.9328** | 40 | 7.0543 | - | - |

(c) Batch_size

| Batch_size | MAE | Batch_size | MAE |
|---|---|---|---|
| 2 | 7.4460 | 32 | 7.4829 |
| 4 | 7.1975 | 64 | 7.0440 |
| 8 | 7.1711 | 128 | 7.0960 |
| **16** | **6.9543** | 256 | 7.0962 |

and the MAE of prediction results is used to evaluate the model performance. First, the length of hidden vector is set to 2, 4, 8, 16, 32, 64, 128 or 256 when other parameters are set at default values. Second, the number of epoch is set to 5, 10, 15, 20, 25, 30, 45 and 50 when the batch_size is set to default and the length of hidden vector is fixed with best values. Third, the batch_size is set to 2, 4, 8, 16, 32, 64, 128, or 256 when other parameters are fixed with best values. The model performance has been summarized in Table 2 and Table 3. The best parameters are bolded.

However, when $N = 0$, $step = 1$, the length of hidden vector is 128 and other hyperparameters are set to the optimized values in Table 2, the MAE of prediction results is 7. 9721. Compared with the best results in Table 2, it is the second best accurate prediction. Finally, to achieve better forecasting performance and assure the hyperparameters for all models, the length of hidden vector is set to 128. The epoch is set to 20 and the batch_size is set to 16. The prediction results are summarized in Table 4.

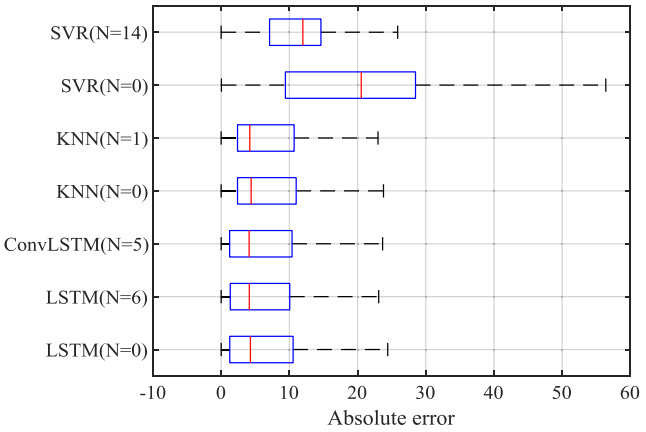



**FIGURE 11.** Boxplot of prediction performance in test data set.

With the lowest MAE value and RMSE value, the results of the best baseline and the best model are highlighted in bold. According to the prediction error analysis, the most accurate prediction occurs when $N$ is set to 6 and *step* is set to 11. Compared with the best result in baseline models, the prediction accuracy is improved by reducing MAE nearly 5.48%. And the RMSE has dropped by 6.23%. The comparison between traffic flow data and predicted values is presented in Figure 10a. Figure 10b and Figure 10c visualize the comparison of predicted traffic flow on a day. It can be seen that the traffic flow predicted by the overall flow and 6 flows is more approximate to the traffic flow data when the traffic flow fluctuates or drops largely in the morning. If $N$ is set to 12, the models always perform better than the baselines when *step* is set to the same value. The prediction results are highlighted in bold with an asterisk in Table 4.

### C. MODEL COMPARISON

The performances of the LSTM model are compared with other prediction approaches which include the Convolutional LSTM (ConvLSTM) model, the KNN model and the SVR model. As a variant of LSTM, it has been applied in handling spatiotemporal data in recent research [42]. The key formulas are show at below:

$$i_t = \sigma(\mathcal{W}_{xi} * \mathcal{X}_t + \mathcal{W}_{hi} * \mathcal{H}_{t-1} + \mathcal{W}_{ci} \odot c_{t-1} + \mathcal{B}_i) \quad (23)$$

$$f_t = \sigma(W_{xf} * \mathcal{X}_t + W_{hf} * \mathcal{H}_{t-1} + \mathcal{W}_{cf} \odot c_{t-1} + \mathcal{B}_f) \quad (24)$$

$$c_t = f_t * c_{t-1} + i_t * \sigma(\mathcal{W}_{xc} * \mathcal{X}_t + \mathcal{W}_{hc} * \mathcal{H}_{t-1} + \mathcal{W}_{hc} \odot c_{t-1} + \mathcal{B}_c) \quad (25)$$

$$o_t = \sigma(\mathcal{W}_{xo} * \mathcal{X}_t + \mathcal{W}_{ho} * \mathcal{H}_{t-1} + \mathcal{W}_{co} \odot c_{t-1} + \mathcal{B}_o) \quad (26)$$

$$\mathcal{H}_t = o_t \odot \tanh(c_{t-1}) \quad (27)$$

where $\mathcal{X}_t \epsilon (N+1, 1)$ and $\mathcal{H}_t \epsilon (N+1, 128) N = 1, 2, \cdots, 14$. The $*$ denotes the convolution operator. Before the $\mathcal{H}_t$ is mapped to the fully connected layer, it is reshaped as a tensor with a size of $(1, 128+128*N)$. In this paper, the ConvLSTM has 128 filters which the size of every filter is 2.

For the SVR model, kernel function is set to the RBF (Radial Basis Function), $C$ is set to 1, and $\varepsilon$ is set to 0.1.

**TABLE 4.** The summary of model results.

(a)

| | *N*=0(Baselines) | | *N*=1 | | *N*=2 | | *N*=3 | | *N*=4 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| *Step*=1 | 11.8671 | 8.1287 | 11.7854 | 7.9004 | 11.7627 | 7.9154 | 11.9510 | 8.1543 | 12.1366 | 8.1747 |
| *Step*=2 | 11.0912 | 7.3487 | 11.0788 | 7.3225 | 11.8467 | 7.9555 | 11.0116 | 7.3399 | 10.9109 | 7.4702 |
| *Step*=3 | 11.0069 | 7.3514 | 12.6008 | 8.4990 | 10.9776 | 7.2498 | 11.2982 | 7.6105 | 10.6821 | 7.2744 |
| *Step*=4 | 11.3260 | 7.9208 | 11.9838 | 7.9167 | 10.9722 | 7.3945 | 12.8491 | 9.8265 | 10.8861 | 7.2019 |
| *Step*=5 | 10.9934 | 7.5225 | 11.2256 | 7.5794 | 11.0640 | 7.3387 | 10.9439 | 7.2319 | 10.7199 | 7.3745 |
| *Step*=6 | 11.0054 | 7.3634 | 10.7527 | 7.2494 | 12.2914 | 8.2474 | 11.1447 | 7.9094 | 10.7746 | 7.2621 |
| *Step*=7 | 10.9259 | 7.4381 | 11.9272 | 8.2765 | 10.7996 | 7.2992 | 10.7972 | 7.2829 | 10.6353 | 7.1406 |
| *Step*=8 | 10.7376 | 7.3746 | 10.6109 | 7.2751 | 10.6307 | 7.1883 | 10.9389 | 7.4134 | 10.6522 | 7.1195 |
| *Step*=9 | 10.8385 | 7.5838 | 10.8429 | 7.3523 | 10.7138 | 7.1089 | 10.4550 | 6.9957 | 10.3601 | 7.1831 |
| *Step*=10 | 10.8864 | 7.6894 | 10.8041 | 7.9031 | 10.5663 | 7.3086 | 10.6226 | 7.2005 | 10.3914 | 7.1961 |
| *Step*=11 | **10.7362** | **7.2359** | 10.2520 | 6.9290 | 10.8970 | 7.2823 | 10.4230 | 7.0664 | 10.8267 | 7.4900 |
| *Step*=12 | 10.7364 | 7.5444 | 10.1754 | 6.8512 | 10.6491 | 7.5237 | 10.5670 | 7.2572 | 10.2515 | 6.9882 |

(b)

| | *N*=5 | | *N*=6 | | *N*=7 | | *N*=8 | | *N*=9 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| *Step*=1 | 11.9441 | 8.6077 | 11.7853 | 7.9534 | 11.8221 | 8.1546 | 12.3062 | 8.3586 | 11.7974 | 7.9698 |
| *Step*=2 | 10.9717 | 7.4447 | 10.9740 | 7.3436 | 11.6007 | 8.1209 | 10.9284 | 7.4229 | 10.9719 | 7.5492 |
| *Step*=3 | 10.8618 | 7.5745 | 11.0036 | 7.7286 | 10.8966 | 7.3399 | 10.5789 | 7.0636 | 10.6578 | 7.2037 |
| *Step*=4 | 10.8364 | 7.4704 | 10.5601 | 7.1129 | 10.6407 | 7.2334 | 12.5817 | 8.3892 | 10.6132 | 7.2283 |
| *Step*=5 | 10.7792 | 7.2586 | 11.2686 | 7.3916 | 11.2418 | 7.7030 | 10.5076 | 7.0184 | 10.5623 | 7.1043 |
| *Step*=6 | 10.8936 | 7.2779 | 10.2437 | 6.8762 | 10.5561 | 6.9876 | 10.8863 | 7.4352 | 10.5416 | 7.0856 |
| *Step*=7 | 10.4734 | 7.1465 | 10.4970 | 7.4847 | 12.1405 | 7.8583 | 10.4074 | 7.1220 | 11.0066 | 7.3829 |
| *Step*=8 | 10.7070 | 7.2776 | 10.5487 | 7.1961 | 10.3536 | 7.2022 | 10.6797 | 7.2846 | 10.3664 | 7.0169 |
| *Step*=9 | 10.9624 | 7.3585 | 10.2053 | 6.9454 | 11.4665 | 7.8270 | 10.7173 | 7.3358 | 10.5999 | 7.3492 |
| *Step*=10 | 10.8812 | 7.4537 | 10.2284 | 7.0470 | 10.4410 | 7.2413 | 10.7058 | 7.2317 | 10.2804 | 6.9800 |
| *Step*=11 | 10.4491 | 7.0627 | **10.0673** | **6.8393** | 10.3873 | 7.0697 | 10.6668 | 7.1871 | 10.2623 | 6.9769 |
| *Step*=12 | 10.5113 | 7.0171 | 10.2260 | 7.0006 | 10.7185 | 7.1460 | 10.2860 | 6.9238 | 10.7549 | 7.4390 |

(c)

| | *N*=10 | | *N*=11 | | *N*=12 | | *N*=13 | | *N*=14 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| *Step*=1 | 11.7805 | 8.2214 | 11.8635 | 7.9876 | **11.7245*** | **7.9608*** | 11.7295 | 7.8918 | 11.8530 | 8.4262 |
| *Step*=2 | 11.0255 | 7.6440 | 10.9825 | 7.3110 | **11.0420*** | **7.2808*** | 11.6362 | 8.0234 | 11.1219 | 7.4142 |
| *Step*=3 | 10.6375 | 7.1538 | 10.6735 | 7.2202 | **10.6399*** | **7.2427*** | 10.7055 | 7.2563 | 11.3291 | 7.4734 |
| *Step*=4 | 10.6666 | 7.1187 | 10.7873 | 7.5452 | **10.8114*** | **7.6381*** | 10.4294 | 7.0511 | 11.0204 | 7.2114 |
| *Step*=5 | 11.5470 | 7.5667 | 10.4665 | 7.0210 | **10.5032*** | **7.0391*** | 10.8454 | 7.1866 | 10.5408 | 7.0828 |
| *Step*=6 | 10.5775 | 7.0014 | 10.9183 | 7.6911 | **10.6092*** | **7.0924*** | 10.4420 | 6.9458 | 10.5395 | 7.0164 |
| *Step*=7 | 11.4393 | 7.4862 | 10.3480 | 7.2812 | **10.8978*** | **7.4369*** | 11.3991 | 7.4247 | 10.4057 | 6.9920 |
| *Step*=8 | 10.7601 | 7.5989 | 10.4612 | 6.9690 | **10.5664*** | **7.1088*** | 10.8098 | 7.5920 | 10.9589 | 7.7774 |
| *Step*=9 | 10.5561 | 7.2952 | 10.3963 | 6.9927 | **10.2485*** | **6.8960*** | 10.4561 | 6.9898 | 10.8278 | 7.2114 |
| *Step*=10 | 10.2913 | 6.9652 | 10.4041 | 7.0131 | **10.3892*** | **7.1557*** | 10.3374 | 6.9318 | 10.7217 | 7.6530 |
| *Step*=11 | 10.3158 | 7.1924 | 11.2490 | 7.6035 | **10.1765*** | **6.9702*** | 10.3551 | 6.9737 | 10.3171 | 6.9000 |
| *Step*=12 | 10.1201 | 6.9596 | 10.5077 | 7.2656 | **10.1731*** | **6.9625*** | 10.4696 | 7.0937 | 10.4224 | 7.2973 |

For the KNN model, the $k$ is set to 5. In the Table 5, best prediction results in different methods have been evaluated by RMSE and MAE. Figure 11 gives the absolute error on test data set predicted by those models.

From Table 5 and Figure 11, it can be seen that split flows improve the performance of traffic flow predictions on various model when the time of training model increased very little. Compared with the result of LSTM model, the prediction results of ConvLSTM model achieve worse performance with more training time. When the information of split flows is not considered, the SVR and the KNN model results have the highest errors.

For the KNN model, the best prediction occurs when $N$ was set to 1 and *step* was set to 8. Compared with the best baseline in the KNN models, the improvements of accuracy are up to 4.66% in MAE and 5.06% in terms of RMSE.

For the SVR model, the evaluations show that the SVR model achieved the best performance when $N$ was set

**TABLE 5.** The best results in different models.

| | Parameter | | Evaluation | | Cost time(s) |
|---|---|---|---|---|---|
| | *N* | *Step* | RMSE | MAE | |
| LSTM | 0 | 11 | 10.7362 | 7.2359 | 31.4898 |
| LSTM | 6 | 11 | 10.0673 | 6.8393 | 32.2892 |
| ConvLSTM | 5 | 11 | 10.2134 | 6.9059 | 129.0560 |
| KNN | 0 | 8 | 11.3281 | 7.8750 | 0.0080 |
| KNN | 1 | 8 | 10.7546 | 7.5078 | 0.0100 |
| SVR | 0 | 2 | 21.4442 | 18.7987 | 0.0300 |
| SVR | 14 | 1 | 14.2697 | 12.0261 | 0.0505 |

**TABLE 6.** List of abbreviations.

| Abbreviation | Full Name |
|---|---|
| ANN | Artificial Neural Network |
| ARIMA | Autoregressive Integrated Moving Average |
| ARIMA-GARCH | Autoregressive Integrated Moving Average with Generalized Autoregressive Conditional Heteroscedasticity |
| BP | Back Propagation Network |
| BPTT | Back Propagation Through Time |
| CLTFP | a novel short-term traffic flow prediction approach based on the combination of Convolutional Neural Network and Long Short-Term Memory<br>C: Convolutional Neural Network<br>L: Long Short-Term Memory<br>T:Traffic<br>F: Flow<br>P:Prediction |
| CNN | Convolutional Neural Network |
| ConvLSTM | Convolutional Long Short-Term Memory |
| DELA | A deep and embedding learning approach that consisted of an embedding component, a Convolutional Neural Networkcomponent and a Long Short-Term Memory component. |
| ETC | Electronic Toll Collection |
| FFNN | Feed Forward Neural Network |
| GRU | Gated Recursive Unit |
| ITS | Intelligent Transportation Systems |
| KNN | K-Nearest Neighbor |
| LSTM | Long Short-Term Memory |
| MAE | Mean Absolute Error |
| MAPE | Mean Absolute Percentage Error |
| MLANN | Multilayer Artificial Neural Network |
| MLP | Multilayer Perceptron |
| RBF | Radial Basis Function |
| RBFN | Radial Basis Function Network |
| RBFNN | Radial Basis Function Neural Networks |
| RMSE | Root Mean Square Error |
| RNN | Recurrent Neural Network |
| RW | Random Walk |
| SAE | Stacked AutoEncoder |
| SARIMA | Seasonal Autoregressive Integrated Moving Average |
| SGD | stochastic gradient descent |
| STFSA | Spatio-temporal feature selection algorithm |
| SVM | Support Vector Machine |
| SVR | Support Vector Regression |
| WNN | Wavelet Neural Network |

to 14 and *step* was set to 1. Compared with the best baseline, the prediction accuracy decreased MAE by 36.03%. The RMSE value drops by 33.46%.

## IV. CONCLUSION

In this paper, we collected the traffic flow data from an expressway exit station in Shanghai and used the LSTM model to predict the overall exit traffic flow. Both the overall flow information and the split flow information from test data were used to forecast the overall flow at later dates. We conducted experiments to analyze the effect of split flow information compared to baseline models, which only considered the overall flow information. We also evaluated other models in the same manner. According to the prediction analysis, the split flow information improves the accuracy of model predictions. We obtained results as below:

(I) When $N$ was set to 6 and *step* was set to 11, the model gave the best prediction performance. The prediction error reduced by 5.48%, as computed by MAE, compared to the baseline. RMSE declined by 6.23%.

(II) The models that considered the information of overall flow and 12 split flows with largest numbers of vehicles always achieved better performances than the baselines.

(III) The proposed method is particularly effective for predicting when large fluctuation occurs.

(IV) LSTM models exhibit better performance than the ConvLSTM models, the KNN models and the SVR models. Incorporating split flow information improves the performance of SVR model and KNN model when the training time increases not very much. For the KNN model, when $N$ was set to 1 and *step* was set to 8, the KNN model performed best. For the SVR model, the results show that if $N$ was set to 14 and *step* was set to 1, the model would achieve best performance.

The results show that the inclusion of split flow information improves overall flow prediction.

This paper is an extension of previous study [36]. Compared with the early version, the improvement is summarized as follows:

(I) Describe the trend of traffic flow accurately.

(II) Explain why 12 split flows should be selected by cross-correlation analysis.

(III) Describe the process of how to determine the values of hyperparameters.

(IV) The comparison of various models shows that the information of split flows makes various models achieve better performance.

## APPENDIX

See Table 6.

## REFERENCES

[1] Z. Zheng, Y. Yang, J. Liu, H.-N. Dai, and Y. Zhang, "Deep and embedded learning approach for traffic flow prediction in urban informatics," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 10, pp. 3927–3939, Oct. 2019, doi: 10.1109/TITS.2019.2909904.

[2] L. Li, L. Qin, X. Qu, J. Zhang, Y. Wang, and B. Ran, "Day-ahead traffic flow forecasting based on a deep belief network optimized by the multi-objective particle swarm algorithm," *Knowl.-Based Syst.*, vol. 172, pp. 1–14, May 2019, doi: 10.1016/j.knosys.2019.01.015.

[3] L. Liu, J. Zhen, G. Li, G. Zhan, Z. He, B. Du, and L. Lin, "Dynamic spatial-temporal representation learning for traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, early access, Jun. 29, 2020, doi: 10.1109/TITS.2020.3002718.

[4] Y.-S. Jeong, Y.-J. Byon, M. M. Castro-Neto, and S. M. Easa, "Supervised weighting-online learning algorithm for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1700–1707, Dec. 2013, doi: 10.1109/TITS.2013.2267735.

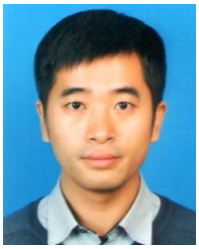
[5] Q. Zhaowei, L. Haitao, L. Zhihui, and Z. Tao, "Short-term traffic flow forecasting method with M-B-LSTM hybrid network," *IEEE Trans. Intell. Transp. Syst.*, early access, Jul. 29, 2020, doi: 10.1109/TITS.2020.3009725.

[6] D. Xu and Y. Shi, "A combined model of random forest and multilayer perceptron to forecast expressway traffic flow," in *Proc. 7th IEEE Int. Conf. Electron. Inf. Emergency Commun. (ICEIEC)*, Jul. 2017, pp. 448–451.
[7] Z. Chen, B. Wu, B. Li, and H. Ruan, "Expressway exit traffic flow prediction for ETC and MTC charging system based on entry traffic flows and LSTM model," *IEEE Access*, vol. 9, pp. 54613–54624, Apr. 2021, doi: 10.1109/ACCESS.2021.3070625.
[8] Y. Lin, L. Li, H. Jing, B. Ran, and D. Sun, "Automated traffic incident detection with a smaller dataset based on generative adversarial networks," *Accident Anal. Prevention*, vol. 144, Sep. 2020, Art. no. 105628, doi: 10.1016/j.aap.2020.105628.
[9] P. Wang, W. Hao, and Y. Jin, "Fine-grained traffic flow prediction of various vehicle types via fusison of multisource data and deep learning approaches," *IEEE Trans. Intell. Transp. Syst.*, early access, Jun. 8, 2020, doi: 10.1109/TITS.2020.2997412.
[10] D. Kang, Y. Lv, and Y.-Y. Chen, "Short-term traffic flow prediction with LSTM recurrent neural network," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–6.
[11] S. Du, T. Li, Y. Yang, X. Gong, and S.-J. Horng, "An LSTM based encoder-decoder model for MultiStep traffic flow prediction," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8.
[12] L. Zhang, N. R. Alharbe, G. Luo, Z. Yao, and Y. Li, "A hybrid forecasting framework based on support vector regression with a modified genetic algorithm and a random forest for traffic flow prediction," *Tsinghua Sci. Technol.*, vol. 23, no. 4, pp. 479–492, Aug. 2018, doi: 10.26599/TST.2018.9010045.
[13] X. Luo, D. Li, Y. Yang, and S. Zhang, "Spatiotemporal traffic flow prediction with KNN and LSTM," *J. Adv. Transp.*, vol. 1, pp. 537–546, Feb. 2019, doi: 10.1155/2019/4145353.
[14] W. Wei, H. Wu, and H. Ma, "An AutoEncoder and LSTM-based traffic flow prediction method," *Sensors*, vol. 19, no. 13, p. 2946, Jul. 2019, doi: 10.3390/s19132946.
[15] S. Suhas, V. V. Kalyan, M. Katti, B. V. A. Prakash, and C. Naveena, "A comprehensive review on traffic prediction for intelligent transport system," in *Proc. Int. Conf. Recent Adv. Electron. Commun. Technol. (ICRAECT)*, Mar. 2017, pp. 138–143.
[16] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proc. 31st Youth Academic Annu. Conf. Chin. Assoc. Autom. (YAC)*, Nov. 2016, pp. 324–328.
[17] H. Dong, L. Jia, X. Sun, C. Li, and Y. Qin, "Road traffic flow prediction with a time-oriented ARIMA model," in *Proc. 5th Int. Joint Conf. (INC, IMS IDC)*, Aug. 2009, pp. 1649–1652.
[18] S. V. Kumar, "Traffic flow prediction using Kalman filtering technique," *Procedia Eng.*, vol. 187, pp. 582–587, Jan. 2017, doi: 10.1016/j.proeng.2017.04.417.
[19] C. Chen, J. Hu, Q. Meng, and Y. Zhang, "Short-time traffic flow prediction with ARIMA-GARCH model," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 607–612.
[20] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *Eur. Transp. Res. Rev.*, vol. 7, no. 3, pp. 1–9, Jun. 2015, doi: 10.1007/s12544-015-0170-8.
[21] H. Zhu, Y. Xie, W. He, C. Sun, K. Zhu, G. Zhou, and N. Ma, "A novel traffic flow forecasting method based on RNN-GCN and BRB," *J. Adv. Transp.*, vol. 2020, pp. 1–11, Oct. 2020, doi: 10.1155/2020/7586154.
[22] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015, doi: 10.1109/TITS.2014.2345663.
[23] J. Zheng, Z. Xin, and C. Zhu, "Traffic volume forecasting based on radial basis function neural network with the consideration of traffic flows at the adjacent intersections," *Transp. Res. C, Emerg. Technol.*, vol. 47, no. 2, pp. 139–154, Jul. 2014, doi: 10.1016/j.trc.2014.06.011.
[24] D. Zeng, J. Xu, J. Gu, L. Liu, and G. Xu, "Short term traffic flow prediction using hybrid ARIMA and ANN models," in *Proc. Workshop Power Electron. Intell. Transp. Syst.*, Aug. 2008, pp. 621–625.
[25] X. Wang, K. An, L. Tang, and X. Chen, "Short term prediction of freeway exiting volume based on SVM and KNN," *Int. J. Transp. Sci. Technol.*, vol. 4, no. 3, pp. 337–352, Sep. 2015, doi: 10.1260/2046-0430.4.3.337.
[26] A. Graves, "Neural networks," in *Supervised Sequence Labelling With Recurrent Neural Networks* 1st. ed. Berlin, Germany: Springer, 2012, pp. 5–35.
[27] L. Li, H. Zhou, H. Liu, C. Zhang, and J. Liu, "A hybrid method coupling empirical mode decomposition and a long short-term memory network to predict missing measured signal data of SHM systems," *Struct. Health Monitor.*, pp. 1–16, Jul. 2020, doi: 10.1177/1475921720932813.
[28] H. Shao and B.-H. Soong, "Traffic flow prediction with long short-term memory networks (LSTMs)," in *Proc. IEEE Region Conf. (TENCON)*, Nov. 2016, pp. 2986–2989.
[29] Y. Tian and L. Pan, "Predicting short-term traffic flow by long short-term memory recurrent neural network," in *Proc. IEEE Int. Conf. Smart City/SocialCom/SustainCom (SmartCity)*, Dec. 2015, pp. 153–158.
[30] S. Wang, J. Zhao, C. Shao, C. Dong, and C. Yin, "Truck traffic flow prediction based on LSTM and GRU methods with sampled GPS data," *IEEE Access*, vol. 8, pp. 208158–208169, 2020, doi: 10.1109/ACCESS.2020.3038788.
[31] Y. Jia, J. Wu, and M. Xu, "Traffic flow prediction with rainfall impact using a deep learning method," *J. Adv. Transp.*, vol. 2017, pp. 1–10, Aug. 2017, doi: 10.1155/2017/6575947.
[32] I. Guyon and A. Elisseeff, "An introduction of variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Jun. 2003, doi: 10.1162/153244303322753616.
[33] Y. Wu and H. Tan, "Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework," 2016, *arXiv:1612.01022*. [Online]. Available: http://arxiv.org/abs/1612.01022
[34] Z. Yan, C. Yu, L. Han, W. Su, and P. Liu, "Short-term traffic flow forecasting method based on CNN+LSTM," *Comput. Eng. Des.*, vol. 40, no. 9, pp. 2620–2624, Sep. 2019.
[35] W. Zhang, Y. Yu, Y. Qi, F. Shu, and Y. Wang, "Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning," *Transportmetrica A, Transp. Sci.*, vol. 15, no. 7, pp. 1–45, Jul. 2019, doi: 10.1080/23249935.2019.1637966.
[36] H. Ruan, B. Wu, B. Li, and Z. Chen, "Highway short-term traffic flow prediction with traffic flows from multi entry stations," in *Proc. 3rd Int. Forum Connected Automated Vehicle Highway Syst. Through China Highway Transp. Soc.*, Dec. 2020, pp. 1–10.
[37] J. Chung, C. K. Gulcehre Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2016, *arXiv: 11412.3555*. [Online]. Available: https://arxiv.org/abs/1412.3555
[38] L. Chen, L. Zheng, J. Yang, D. Xia, and W. Liu, "Short-term traffic flow prediction: From the perspective of traffic flow decomposition," *Neurocomputing*, vol. 413, pp. 444–456, Nov. 2020, doi: 10.1016/j.neucom.2020.07.009.
[39] Y. Yu, Y. Zhang, S. Qian, S. Wang, Y. Hu, and B. Yin, "A low rank dynamic mode decomposition model for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, early access, May 27, 2020, doi: 10.1109/TITS.2020.2997412.
[40] D. Ma, X. Song, and P. Li, "Daily traffic flow forecasting through a contextual convolutional recurrent neural network modeling inter- and intra-day traffic patterns," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 2627–2636, May 2021, doi: 10.1109/TITS.2020.2973279.
[41] *Google Map*. Accessed: Nov. 2021. [Online]. Available: https://www.google.com.hk/maps/@31.1253106,121.8174882,9.75z?hl=en
[42] H. Zheng, F. Lin, X. Feng, and Y. Chen, "A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, early access, Jun. 9, 2020, doi: 10.1109/TITS.2020.2997352.

**HOUSONG RUAN** received the B.S. degree from the Anhui University of Finance and Economics, China, in 2019. He is currently pursuing the master's degree with Xi'an Jiaotong University. His research interests include traffic flow forecast and seismic wave inversion.

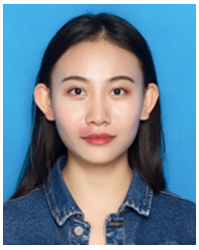
**BANGYU WU** (Member, IEEE) received the B.S. degree in information engineering and the Ph.D. degree in information and communication engineering from Xi'an Jiaotong University, Xi'an, China, in 2005 and 2012, respectively.

From 2007 to 2011, he was a Visiting Scholar with the University of California Santa Cruz. He also worked as a Senior Geophysicist with Statoil (Beijing) Technology Service Company Ltd., from 2012 to 2015. He is currently an Associate Professor with the School of Mathematics and Statistics, Xi'an Jiaotong University. His research interests include seismic wave modeling and migration/inversion, signal processing, and machine learning.



**BIN LI** was born in Huanglong, Yan'an, Shaanxi, China, in 1982. He received the B.S., M.S., and Ph.D. degrees in automobile engineering from Chang'an University, in 2005, 2008, and 2013, respectively.

From 2014 to 2019, he was a Lecturer with the School of Automobile, Chang'an University. Since 2020, he has been the Vice Chairman of the Department of Automobile Engineering, Chang'an University. He is the coauthor of six books and more than 15 articles. His research interests include cooperative vehicle infrastructure systems, optimization design technology of electric vehicle, and traffic big data analysis. He is the member of working committee of automated driving of China.



**ZHU CHEN** received the B.S. degree with the School of Mathematical Sciences from Guizhou Normal University, Guiyang, China, in 2019. She is currently pursuing the master's degree with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China.

Her research interest includes time series forecast.

**WEIGUO YUN** was born in Minqin, Wuwei, Gansu, China, in 1981. He received the B.S., M.S., and Ph.D. degrees in automobile engineering from Chang'an University, in 2005, 2008, and 2018., respectively

From 2008 to 2014, he worked with Shan'xi Heavy Duty Automobile Ltd. Since 2010, he has been the Director of the Department of Automobile NVH Performance Development, Shan'xi Heavy Duty Automobile Company. Since 2015, he has been working with Geely Automobile Group and mainly responsible for new energy vehicle verification and prototype cars manufacturing. Since he has finished more than six commercial vehicles testing and building. He is the author of more than 15 articles. His research interests include testing technology of vehicle, optimization design technology of electric vehicle, and NVH technology. He is the Expert Member of SAE, China.