

Article

A Novel Hybrid Model for Short-Term Traffic Flow Prediction Based on Extreme Learning Machine and Improved Kernel Density Estimation

Leina Zhao ^{1,2}, Yujia Bai ^{1,*}, Sishi Zhang ¹, Yanpeng Wang ² , Jie Kang ³ and Wenxuan Zhang ³¹ College of Mathematics and Statistics, Chongqing Jiaotong University, Chongqing 400074, China² School of Traffic and Transportation, Chongqing Jiaotong University, Chongqing 400074, China³ School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China

* Correspondence: 622200150020@mails.cqjtu.edu.cn; Tel.: +86-18715865100

Abstract: Short-term traffic flow prediction is the basis of and ensures intelligent traffic control. However, the conventional models cannot make accurate predictions due to the strong nonlinearity and randomness in short-term traffic flow data. To this end, the authors of this paper developed a novel hybrid model based on extreme learning machine (ELM), adaptive kernel density estimation (AKDE), and conditional kernel density estimation (CKDE). Specifically, the ELM model was employed for nonlinear prediction. Then, AKDE was established to estimate the bandwidth of CKDE (i.e., AKDE-CKDE), which predicted the training residuals obtained by ELM. Finally, the predicted results of the two models were superimposed to derive the final prediction of the hybrid model. Two case studies based on measured data were conducted to evaluate the performance of the proposed method. The experimental results indicate that the proposed method can realize a significant improvement in terms of forecasting accuracy in comparison with the other concerned models. For instance, it performed better than the single ELM model, with an improvement in the evaluation criterion of a mean relative percentage error of 7.46%.

Keywords: short-term traffic flow prediction; hybrid model; ELM; AKDE-CKDE



Citation: Zhao, L.; Bai, Y.; Zhang, S.; Wang, Y.; Kang, J.; Zhang, W. A Novel Hybrid Model for Short-Term Traffic Flow Prediction Based on Extreme Learning Machine and Improved Kernel Density Estimation. *Sustainability* **2022**, *14*, 16361. <https://doi.org/10.3390/su142416361>

Academic Editors: Qixiu Cheng, Kai Huang, Yongxiang Zhang and Yu Yao

Received: 8 November 2022

Accepted: 5 December 2022

Published: 7 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the acceleration of urbanization and the rapid increase in car ownership, traffic congestion in urban areas is becoming more and more serious, leading to a series of social problems, such as traffic accidents, air pollution, energy waste, and so on. These problems have greatly decreased the living standard of human beings. The emergence of the intelligent transportation system (ITS) has effectively alleviated traffic congestion and traffic accidents, thereby improving the efficiency of urban traffic operations and reducing environmental pollution [1].

Short-term traffic flow prediction is one of the crucial tasks of ITS. It aims to forecast the variation in traffic flow soon from a few seconds to a few hours based on historical traffic data. The accuracy and efficiency of prediction play a decisive role in the performance of path guidance and transportation management [2]. In recent decades, to enhance prediction accuracy, domestic and foreign scholars have put forward a wide variety of approaches. Generally, research on traffic flow prediction falls into the following three categories: statistical theoretical models, intelligent models, and hybrid models.

The commonly used statistical models include time-series models (e.g., autoregressive integrated moving average (ARIMA), seasonal ARIMA (SARIMA), etc.) [3–6], the Kalman filtering model [7,8], the hidden Markov model [9], etc. All of them can obtain linear characteristics hidden in traffic flow data by selecting appropriate parameters. In general, the statistical theoretical models may be more suitable for short-term forecasting and widely utilized in practice due to the simpler model structure and the lower requirement for

data [10]. However, related studies have clarified that these models ignore the interferences of random factors in traffic flow and cannot explain the short-term traffic flow data with strong nonstationarity and high nonlinearity [11–13].

Due to the rapid development of computer technology, a considerable number of prediction technologies based on intelligent models have been developed to achieve high-precision predictions for traffic series with nonlinear and nonstationary characteristics. Commonly used intelligent prediction models in short-term traffic flow prediction include the wavelet neural network (WNN) [14,15], support vector machine (SVM) [16,17], least square SVM (LSSVM) [11], long short-term memory (LSTM) [18,19], and so on. These models can cope with the nonlinear component of the input signal effectively and achieve more accurate prediction results than statistical theoretical models. In addition, it is worth mentioning the ELM model. Huang et al. [20] proposed a new learning algorithm called extreme learning machine (ELM), which exhibits a good generalization performance for feedforward neural networks. This method is different from the previous feedforward neural network parameter adjustment method, which does not need to adjust parameters iteratively, and has the advantages of low complexity and fast convergence. However, the over-fitting problem cannot be addressed completely in ELM, which could affect the model's prediction accuracy [21]. For this reason, many scholars have proposed a variety of improved ELM applications, which have achieved satisfactory performance. For example, Cai et al. [22] and Cui et al. [23] proposed a new PSO-ELM model based on particle swarm optimization, an extreme learning machine (GSA-ELM) optimized by the gravity search algorithm to predict short-term traffic flow. The prediction results indicated that both models mentioned above could effectively improve the prediction accuracy compared to the standalone basic ELM model. Obviously, the aforementioned single models have advantages and high adaptability to nonlinear data. Nevertheless, it is hard to obtain satisfactory predictions using only a single prediction model to predict traffic flow with nonlinear, nonstationary, and random characteristics. At the same time, the standalone basic ELM has certain limitations in practice, which require further research.

Considering the defects of single models, hybrid models based on probabilistic characteristics have been developed and have become increasingly popular in recent years because they can explain the random components of short-term traffic flow data. Recently, numerous scholars have combined the deterministic prediction model with the probabilistic prediction model to establish hybrid prediction models, which capture the nonlinear and random characteristics embedded in the traffic flow time series, respectively. At the same time, the hybrid models can give full play to the strengths of each model to provide more satisfactory predictions. So, a variety of probability estimation approaches for predicting have been proposed. The typical representative of the probabilistic model is Gaussian process regression (GPR), widely used for short-term traffic flow prediction. For example, hybrid models based on the GPR model were established to forecast short-term traffic flow and gained satisfactory prediction performance in Refs. [24,25]. In addition, kernel density estimation (KDE) as a nonparametric model can also provide a probabilistic prediction. The most striking feature of KDE is that it can directly use sample data without any parameter assumptions to estimate the target object. Zhou et al. [26] combined the k-means LSTM network model and nonparametric KDE method with bandwidth optimization for wind energy prediction. The results showed that the proposed model had higher prediction accuracy. Jeon et al. [27] used conditional KDE (CKDE) and Monte Carlo simulation of a statistical model for wind power density forecasting and generated satisfactory prediction results. To the best of our knowledge, the related research of the KDE model is rarely involved in the field of traffic flow prediction.

Generally, as an improvement of KDE, CKDE can fully use prior information. Meanwhile, the determination of the kernel bandwidth is a crucial step in the KDE method [28]. Nevertheless, the overall optimal bandwidth of CKDE cannot adjust according to the local interval's data density, resulting in poor local adaptability. In order to tackle this problem, the adaptive KDE (AKDE) method was employed to improve the CKDE method and

enhance its local adaptability [29]. Furthermore, the high complexity of traffic systems leads to significant randomness of traffic flow, and considering stochastic factors ensures accurate traffic flow prediction. Probabilistic density estimation can effectively quantify the uncertainty of traffic flow and provide more comprehensive information for traffic flow prediction. In light of the above, it is essential to develop a novel effective prediction method further to improve the accuracy of short-term traffic flow prediction. The authors of this paper developed an innovative hybrid short-term traffic flow forecasting model based on the ELM, AKDE, and CKDE. Specifically, the ELM model was adopted to predict the original traffic flow sequence, and the training residuals were obtained. Secondly, AKDE was utilized to estimate the variance in each dimension of the reconstructed samples. Then, the variance was used to replace the relevant parameters of CKDE, and the AKDE-CKDE model was established to forecast the residuals. Thirdly, the final prediction results were obtained by summing up the prediction values of the ELM and AKDE-CKDE model. Finally, the proposed model was analyzed based on two groups of traffic flow data. In order to better exhibit the performance of the proposed model, the authors selected the ARIMA, LSSVM, ELM-CKDE, ELM, and CKDE methods for comparison. Some conclusions are drawn in the end.

The main contributions of the proposed model are:

- A novel hybrid predictor based on the ELM, AKDE, and CKDE is proposed for short-term traffic flow prediction. The main characteristic of the predictor is that it considers the nonlinearity and randomness characteristics of traffic flow data, making it more suitable for the actual situation;
- The corresponding parameters of CKDE are replaced by the variance in the reconstructed residual samples estimated by AKDE, which improves the model's adaptability. In addition, AKDE-CKDE can directly use the sample data for distribution estimation without any parameter assumptions;
- Through extensive experiments on two real-world datasets at the intersection of the main road in the main urban area of Chongqing, the results show that the proposed hybrid model can increase the precision of urban road traffic flow prediction.

The remainder of this article is organized as follows. The basic principles of the methods and the hybrid forecasting model are briefly introduced in Section 2. Two case studies were conducted based on actual traffic flow data, and the corresponding results and analysis are given in Section 3. Finally, Section 4 summarizes some of the main conclusions.

2. Materials and Methods

The ELM, as a single hidden layer feedforward neural network, can randomly initialize the weights and thresholds of the input layer and hidden layer and get the corresponding output weights. It has the advantages of fewer training parameters, faster learning speed, as well as better generalization performance [30]. On the other hand, the combination of AKDE and CKDE can effectively and quickly obtain the probability density function (PDF) of the target variable. The authors of this paper combined the merits of these two models and constructed a new hybrid prediction model, i.e., the ELM-AKDE-CKDE.

2.1. Extreme Learning Machine (ELM)

As an improved single hidden layer feedforward neural network, the ELM has the capacity of training samples without resetting the cost and threshold value, and the optimal connection and bias parameters can be obtained by solving the matrix equation [20]. A typical single hidden layer feedforward neural network is shown in Figure 1.

Suppose there are n arbitrary training samples $\{x_i, y_i\}$, $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$, $y_i = [y_{i1}, y_{i2}, \dots, y_{im}]^T \in R^m$. w_{ij} ($i = 1, \dots, n$, $j = 1, \dots, l$) is the connection weight between the input layer and the hidden layer; β_{jk} ($j = 1, \dots, l$, $k = 1, \dots, m$) denotes the

connection weight between the hidden layer and the output layer; $b_k (k = 1, \dots, l)$ is the hidden layer bias value. Then, the ELM model can be formulated as:

$$t_i = \sum_{j=1}^l \beta_j g(w_j x_i + b_j), \quad i = 1, \dots, n \quad (1)$$

where $g(x)$ is the activation function; $\beta_j = [\beta_{j1}, \beta_{j2}, \dots, \beta_{jm}]^T$, $w_j = [w_{j1}, w_{j2}, \dots, w_{jn}]^T$.

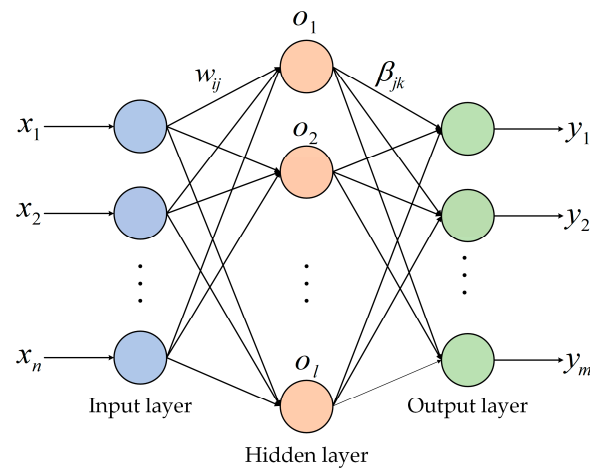


Figure 1. ELM architecture.

Its matrix form is:

$$H\beta = T \quad (2)$$

where H is the output matrix of the hidden layer; β is the matrix of the output weights; T is the output vector.

$$H = \begin{bmatrix} g(w_1 x_1 + b_1) & g(w_2 x_1 + b_2) & \cdots & g(w_l x_1 + b_l) \\ g(w_1 x_2 + b_1) & g(w_2 x_2 + b_2) & \cdots & g(w_l x_2 + b_l) \\ \vdots & \vdots & \ddots & \vdots \\ g(w_1 x_n + b_1) & g(w_2 x_n + b_2) & \cdots & g(w_l x_n + b_l) \end{bmatrix}_{n \times l} \quad (3)$$

$$\beta = [\beta_1^T \quad \beta_2^T \quad \cdots \quad \beta_l^T]_{l \times m} \quad (4)$$

$$T = [t_1^T \quad t_2^T \quad \cdots \quad t_m^T]_{n \times m} \quad (5)$$

The goal of network learning is to minimize the output error of the neural network, i.e.,

$$\sum_{j=1}^n \|t_j - y_j\| \rightarrow 0 \quad (6)$$

By training the single hidden layer neural network to obtain optimal $\hat{\beta}$, which is calculated as:

$$\|H\hat{\beta} - T\| = \min_{\beta} \|H\beta - T\| \quad (7)$$

The elimination of $\hat{\beta}$ gives

$$\hat{\beta} = H^+ T \quad (8)$$

where H^+ is the generalized inverse of matrix H .

The ELM algorithm can be summarized as follows:

Given a training set $\{(x_i, y_i) | x_i \in R^n, y_i \in R^m, i = 1, 2, \dots, n\}$:

1. Determine the specific structure of the ELM network, such as the hidden neuron node number l and the hidden layer activation function $g(x)$;
2. Randomly determine the input weight w_{ij} ($i = 1, \dots, n, j = 1, \dots, l$) and bias b_k ($k = 1, \dots, l$) of the hidden neuron;
3. Calculate the hidden layer output matrix H in Equation (3);
4. Calculate the output weight $\hat{\beta}$ in Equation (8).

2.2. Adaptive Kernel Density Estimation and Conditional KDE (AKDE-CKDE)

As a matter of fact, the choice of bandwidth matrix has a great effect on the estimation results, while the selection of the kernel function may have a minor effect [10]. Therefore, the authors adopted the AKDE method selected by plug-in bandwidth, which can effectively and quickly obtain the probability density estimation function [29]. After that, the coefficients of CKDE were estimated by the variance obtained from the probability density of the sample data in the AKDE method. The detailed illustration of AKDE-CKDE in traffic flow prediction is shown as follows:

Assume that a set of discrete time series of traffic flow after data processing is $\{x_1, x_2, \dots, x_n\}$. For one-step ahead prediction, N d -dimensional explanatory variables $x_t = [x(t), x(t+1), \dots, x(t+d-1)]$ and N target variables $y_t = [x(t+d)]$, $t = 1, 2, \dots, N$ can be constructed by the following equation:

$$\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} x(1) & \cdots & x(d) \\ \vdots & \vdots & \vdots \\ x(N) & \cdots & x(n-1) \end{bmatrix}; \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} x(d+1) \\ \vdots \\ x(n) \end{bmatrix} \quad (9)$$

where $N = n - d$. Then, the sets of x_t and y_t ($t = 1, 2, \dots, N$) can be regarded as independent samples of random vector x ($x \in R^d$) and random variable y ($y \in R$), respectively. Combine x and y , a random vector $z = (x, y) \in R^{d+1}$ with the sample $\{z_t = (x_t, y_t)\}$ can be constructed.

Then, the multi-dimensional kernel function of the random vector z is shown as

$$\begin{aligned} \hat{f}(z) &= \hat{f}(x, y) = \frac{1}{N|B_z|} \sum_{t=1}^N K_{d+1}[B_z^{-1}(z - z_t)] \\ &= \frac{1}{N|B_x| \cdot |B_y|} \sum_{t=1}^N \left\{ K_d[B_x^{-1}(x - x_t)] \cdot K[B_y^{-1}(y - y_t)] \right\} \end{aligned} \quad (10)$$

Similarly, the multi-dimensional kernel density estimation for x is given by

$$\hat{f}(x) = \frac{1}{N|B_x|} \sum_{t=1}^N K_d[B_x^{-1}(x - x_t)] \quad (11)$$

where $K_d(\cdot)$ denotes Gaussian kernel density function. The Gaussian kernel function is often used as the kernel function due to its advantages of simplicity of use. Its expression is shown as

$$\begin{aligned} K_{d+1}(u) &= \left(\frac{1}{\sqrt{2\pi}} \right)^{d+1} \cdot \prod_{i=1}^{d+1} \exp\left(-\frac{u_i^2}{2}\right) \\ u &= (u_1, u_2, \dots, u_{d+1})^T \end{aligned} \quad (12)$$

On the other hand, B_z represents a symmetric and positive definite kernel bandwidth matrix. For simplicity, the diagonal matrix is used as the kernel bandwidth matrix in this paper, and its expression is

$$B_z = \begin{bmatrix} b_1 & & & \\ & b_2 & & \\ & & \ddots & \\ & & & b_{d+1} \end{bmatrix} \quad (13)$$

where b_1, b_2, \dots, b_d are the bandwidth parameters corresponding to each dimension of the independent variable x , b_{d+1} is the bandwidth parameter of y . b_1, b_2, \dots, b_d and b_{d+1} determine the smoothness in the x -direction and y -direction, respectively.

Then, in this study, adaptive kernel density estimation via diffusion was utilized to obtain the mean and variance of the grid points [29], which are shown as

$$E_i(x) = \sum_{j=1}^{\lambda} P_{ij} \cdot x_{ij} \quad (14)$$

$$\sigma_i^2 = E_i(x^2) - (E_i(x))^2, \quad i = 1, 2, \dots, d + 1 \quad (15)$$

where $j = 1, 2, \dots, \lambda$ denotes the number of discrete grid points, which is large enough. In this study, normal reference criterion (NRC) was employed to determine the value of bandwidth parameter b_i ($i = 1, 2, \dots, d + 1$) [31]. Both of them can be calculated by:

$$b_i = \left(\frac{4}{(d+2)(n-d)} \right)^{(1/d+4)} \cdot \sigma_i, \quad (i = 1, 2, \dots, d)$$

$$b_y = \left(\frac{4}{(d+2)(n-d)} \right)^{(1/d+4)} \cdot \sigma_y, \quad (y = d + 1) \quad (16)$$

where σ_i ($i = 1, 2, \dots, d + 1$) is the standard deviation of the grid point probability density.

Based on the above results, the distribution of target variable y under the condition of explanatory variable x can be expressed as

$$\hat{f}(y|x) = \frac{\hat{f}(z)}{\hat{f}(x)} = \sum_{t=1}^N \left\{ w_t(x) \cdot \frac{1}{|B_y|} \cdot K[B_y^{-1}(y - y_t)] \right\}$$

$$w_t(x) = \frac{K_d[B_x^{-1}(x - x_t)]}{\sum_{t=1}^N K_d[B_x^{-1}(x - x_t)]} \quad (17)$$

The conditional expectation and variance of y can be calculated by utilizing Equation (17), which are shown as

$$\mu = \int y \hat{f}(y|x) dy = \sum_{t=1}^N w_t(x) \cdot y_t \quad (18)$$

$$\sigma^2 = \int (y - \mu)^2 \cdot \hat{f}(y|x) dy = |B_y|^2 + \sum_{t=1}^N w_t(x) \cdot y_t^2 - \mu^2 \quad (19)$$

In this way, the one-step ahead forecasting results can be produced by

$$\hat{x}(n+1) = \hat{\mu}(n+1) = \sum_{t=1}^N w_t(x_{N+1}) \cdot y_t \quad (20)$$

$$\hat{\sigma}^2(n+1) = |B_y|^2 + \sum_{t=1}^N w_t(x_{N+1}) \cdot y_t^2 - \mu^2 \quad (21)$$

$$\hat{f}(n+1) = \sum_{t=1}^N \left\{ w_t(x_{N+1}) \cdot \frac{1}{|B_y|} \cdot K[B_y^{-1}(y - y_t)] \right\} \quad (22)$$

where $\hat{x}(n+1)$, $\hat{\sigma}^2(n+1)$ and $\hat{f}(n+1)$ denote the one-step-ahead predicted value, variance and PDF at the time $n+1$, respectively.

2.3. Hybrid Forecasting Model

Through the above brief review, a novel hybrid model-ELM-AKDE-CKDE was employed to enhance prediction accuracy. The ELM model was applied to predict the short-term traffic flow and the training residuals were obtained. Then, the variance of each dimension of the reconstructed sample estimated by AKDE was used to substitute the corresponding parameters of the CKDE model and obtain a one-step-ahead estimation. At the final forecasting task, the prediction result was obtained by the superposition of the two models' predicted values. The specific flowchart of the ELM-AKDE-CKDE model is shown in Figure 2, and the complete steps are shown as follows:

1. Divide original data into two parts, including the training part $\{x(1), \dots, x(n)\}$ and the forecasting part $\{x(n+1), \dots, x(n+N)\}$;
2. Establish the ELM network, and set the hidden node number l and hidden node output function $g(x)$, by which the prediction results $\{\hat{x}'(n+1), \dots, \hat{x}'(n+N)\}$ and training residuals $\{r(1), \dots, r(n)\}$ can be obtained;
3. Replace the corresponding parameters of CKDE with the variance in the reconstructed residual samples estimated by AKDE, then implement one-step-ahead estimation for the residual sequences $\{r(1), \dots, r(n)\}$, by which the predictive value of the $n+1$ th residual data $\hat{r}(n+1)$ can be estimated by AKDE-CKDE;
4. Update the training part to $\{x(2), \dots, x(n+1)\}$ and repeat steps 2–3, and the corresponding residual forecasting result $\hat{r}(n+2)$ can be obtained. Continue one-step ahead prediction until the overall forecasting part is predicted, and the predicted values of the training residuals $\{\hat{r}(n+1), \dots, \hat{r}(n+N)\}$ can be obtained;
5. Summarize the predicted result of ELM $\hat{x}'(n+1)$ and the predicted result of AKDE-CKDE $\hat{r}(n+1)$ and gain the ultimate prediction results $\hat{x}(n+1)$, i.e., $\hat{x}(n+1) = \hat{x}'(n+1) + \hat{r}(n+1)$. By analogy, the final forecasting results $\{\hat{x}(n+1), \dots, \hat{x}(n+N)\}$ can be obtained;
6. Analyze the forecasting results and evaluate the performance of the proposed model via comparing it with the involved models.

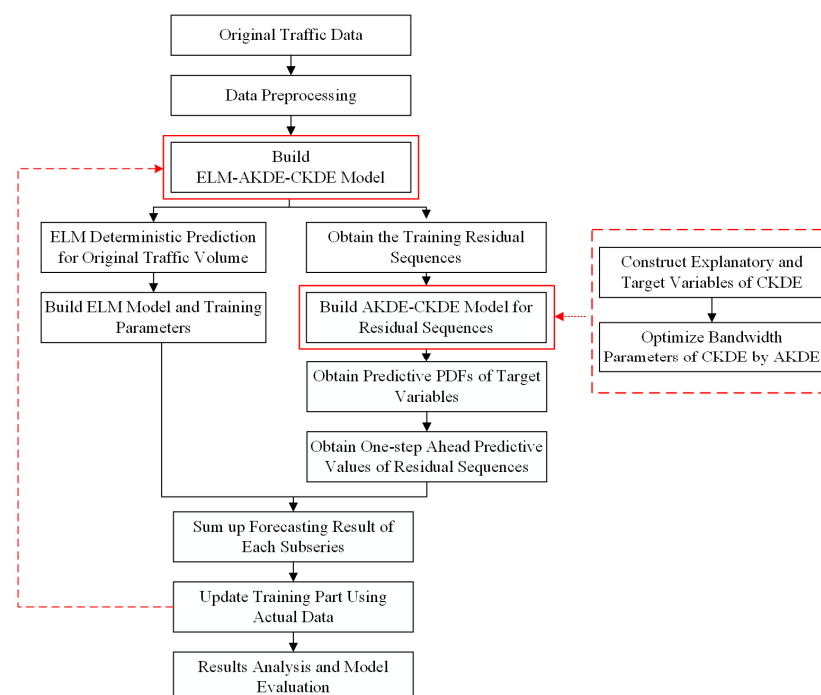


Figure 2. Flowchart of the proposed model.

3. Case Study

3.1. Data Description

To better present the performance of the proposed model, two groups of collected data were utilized for prediction. Dataset 1 and dataset 2 came from the A and B intersections of the main road in the main urban area of Chongqing, respectively, as shown in Figure 3. The collection lasted for a week with a statistical interval of 5 min. A total of 2016 sample data were contained in each dataset.



Figure 3. Location of the intersection in Chongqing.

In this section, the predictive performance of the proposed model based on dataset 1 is presented first. For the sake of making the prediction results more convincing, two-thirds of the data were used to construct and train the model, and the rest were utilized to evaluate the performance [32]. The statistical results of dataset 1 are shown in Figure 4 and Table 1. It should be noted that skewness = 0 and kurtosis = 3 mean that these data overall obey Gaussian distribution, and a value farther away from the target value indicates a stronger non-Gaussianity characteristic. In Table 1, it can be seen that dataset 1 fluctuates severely and has strong nonstationarity and non-Gaussianity.

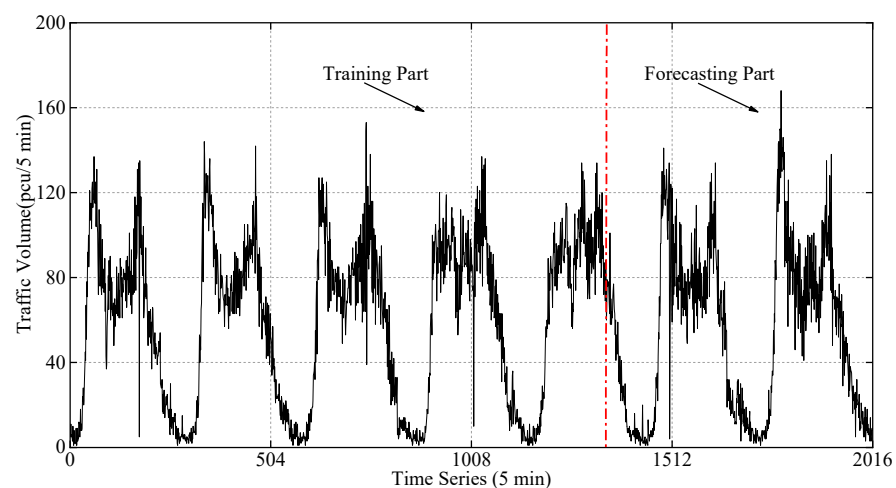


Figure 4. Traffic volume time series (dataset 1).

Table 1. The statistical information of the dataset 1.

Data Source	Mean	Std.	Maximum	Minimum	Skewness	Kurtosis
Dataset 1	57.9851	38.5088	168	1	−0.007	−1.172

3.2. Evaluation Criteria

In order to quantitatively evaluate the performance of the proposed model for short-term traffic flow, the following four frequently used indicators were selected in this study: mean absolute error (*MAE*), mean relative percentage error (*MRPE*), root mean square error (*RMSE*), and root mean square relative error (*RMSRE*). Their specific mathematical expressions are displayed as follows:

The *MAE* represents the mean of the absolute error between the predicted and measured value:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (23)$$

The *MRPE* was used to measure the relative errors between the average predicted value and real value on the test set:

$$MRPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (24)$$

The average differences between the measurements and the predictive values of the method were measured by *RMSE*:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (25)$$

The *RMSRE* represents the standard deviation in the relative error of the prediction:

$$RMSRE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2} \quad (26)$$

where y_i and \hat{y}_i represent the measured value and predicted value, respectively. It is obviously seen in Equations (23)–(26) that the smaller the values of *MAE*, *MRPE*, *RMSE*, and *RMSRE*, the higher the prediction accuracy.

3.3. Performance Evaluation

For the sake of reflecting the superiority of the proposed model, five other models, including the ARIMA model, LSSVM model, ELM-CKDE model, ELM model, and CKDE model, were employed for comparisons. In fact, different parameters may have great impacts on the performance of the prediction method. The ARIMA model, as a statistical model most commonly used for time-series forecasting, can well capture the linear relationships in short-duration traffic volume data [3]. Here, we used the ARIMA (1,1,1) model to predict the traffic flow. As for the ELM method, we set 30 neuron nodes for the hidden layer and generated randomly input weight and bias. For the CKDE nonparametric method, the sample data can be used directly to estimate the distribution without any parameter assumptions. It should be noted that, all experiments were run with the aid of MATLAB 2019a software on a 2.40 GHz PC with I5-1135G7 and 16 GB RAM. In addition, each method was run 10 times independently to mitigate the influence of randomness. On this basis, short-term traffic flow prediction was implemented, and the corresponding forecasting results are given below.

3.4. Traffic Flow Prediction

Taking the case study based on dataset 1 as an example, the prediction process of the proposed method is briefly explained below. In this study, the sigmoid function was chosen as the hidden layer activation function of the ELM network, and its mathematical expression is shown as

$$g(x) = \frac{1}{1 + e^{-x}} \quad (27)$$

According to the above parameter settings, the ELM-AKDE-CKDE model was constructed for experiments. We manually adjusted the parameter value that denotes the input vector dimension of the ELM model mentioned above and compared the *MRPEs* of the prediction results under different dimensional values, as shown in Figure 5. When the dimension of the input vector was 9, the prediction result achieved the best *MRPE*. This means that each input vector of the prediction model was composed of nine consecutive data values in the original traffic flow data, and the corresponding output value is the predicted value of the tenth data point after the initial nine. Finally, we select the 9-dimensional input vector for the experiments to obtain the training residuals of the ELM network.

One-step ahead traffic flow prediction was adopted to illustrate the performance of the proposed method. After subseries reconstruction was achieved, the matrix B_z in CKDE was determined by employing the AKDE and NRC methods. Finally, the estimated probability distribution of traffic flow for the ELM-AKDE-CKDE model prediction results was obtained. Analogously, the probabilistic prediction results could be obtained by applying other probabilistic estimation models, and the final one-step-ahead results, including the predictive PDF and single-value prediction, were generated and are shown in Figure 6. The above procedure was executed for the other training data, and the corresponding prediction results are provided.

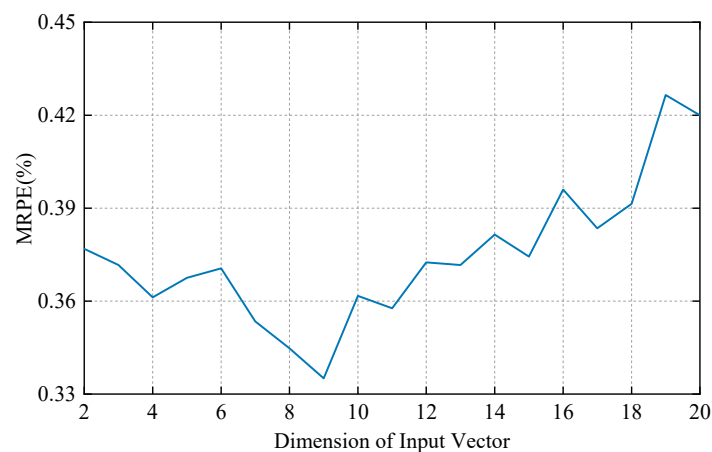


Figure 5. The *MRPEs* of the prediction results by different dimensions of input vector (dataset 1).

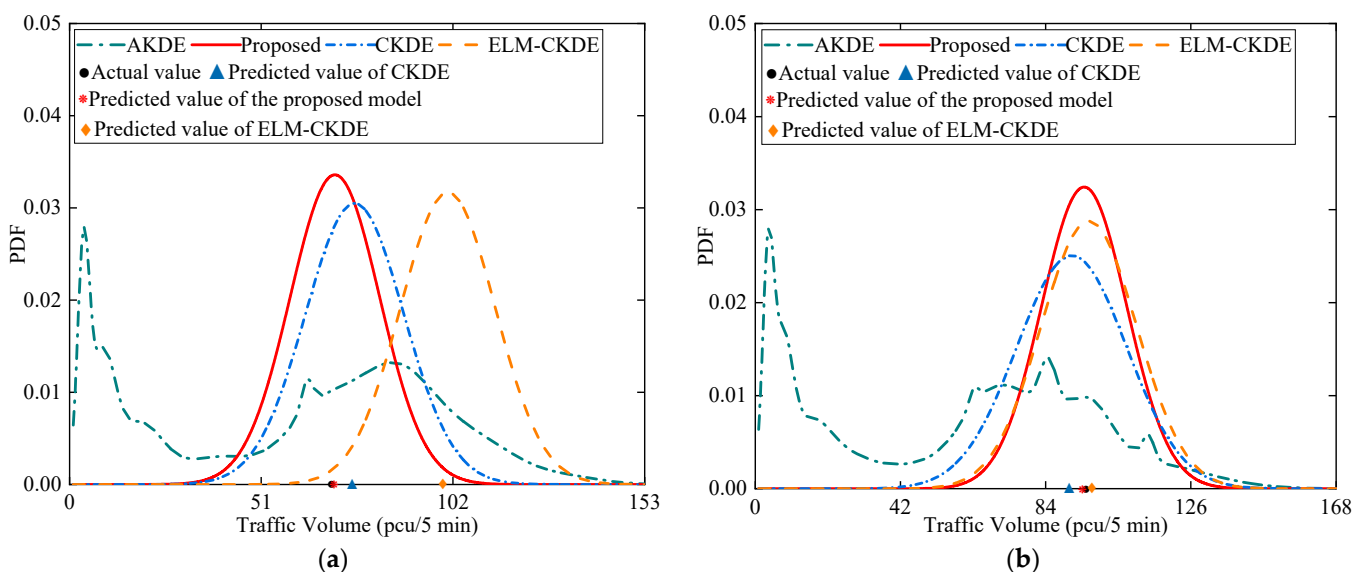


Figure 6. Predictive PDFs, predictive values and actual values of the 14th (a) and the 560th (b) data points in the test part of dataset 1.

3.5. Prediction Results and Analysis

After obtaining the measured data and predicted data, the evaluation index values of the five models involved were calculated, as shown in Table 2. Compared to the other models, the decreased percentage of the proposed model in this paper is shown in Table 3. For the simplicity of the description, the proposed model is abbreviated as proposed. The results can be seen in Tables 2 and 3.

Table 2. Result comparisons of different models (dataset 1).

Model	MAE	MRPE	RMSE	RMSRE
Proposed	9.174	0.335	13.534	1.210%
ARIMA	9.359	0.345	13.751	1.289%
ELM-CKDE	9.890	0.370	14.774	1.248%
ELM	9.274	0.362	13.664	1.273%
CKDE	9.583	0.389	14.175	1.311%
LSSVM	9.395	0.342	13.700	1.257%

Table 3. Improved percentages of the other models by the proposed method (dataset 1).

	MAE (%)	MRPE (%)	RMSE (%)	RMSRE (%)
ARIMA vs. proposed	1.98	2.90	1.58	6.13
ELM-CKDE vs. proposed	7.24	9.46	8.39	3.04
ELM vs. proposed	1.08	7.46	0.95	4.95
CKDE vs. proposed	4.27	13.88	4.52	7.70
LSSVM vs. proposed	2.35	2.05	1.21	3.74

In terms of the single models, two nonlinear models, including ELM and LSSVM, achieved better prediction than the other models. The reason could be that the nonlinear information in dataset 1 is more significant than the linear information. In other words, the ELM and LSSVM models focus on addressing the problem of nonlinear classification and prediction, and they thus outperformed the traditional statistical method ARIMA, but the accuracy was still low. Although CKDE takes the stochastic characteristics of the data into account, it performed the worst overall, and the reason could be that the linear and nonlinear components of the data were ignored when the individual CKDE model predicted traffic flow.

Compared to the single models, the results show that the proposed model produced overall improvements in the experiment, and the reason could be attributed to the fact that the combination of the ELM and AKDE-CKDE could not only extract multiple characteristics embedded in the data but also utilize the strengths of the individual models.

Based on the comparisons between the ELM-AKDE-CKDE and ELM-CKDE models, the improvements by the proposed method in terms of MAE, RMSE, MRPE, and RMSRE were 7.24%, 9.46%, 8.39%, and 3.04%, respectively. In addition, a similar comparison was conducted between the ELM and the proposed model. The results indicate that the AKDE-CKDE method surpassed CKDE in boosting forecasting accuracy; the reason may be that AKDE-CKDE is more effective in dealing with the data randomness of the residuals obtained by ELM. Because AKDE usually has the obvious advantages of being more adaptive, and the overall optimal bandwidth can be adjusted according to the sample density of the characteristic variable data.

In order to more intuitively compare the prediction results of the proposed model with other involved models, Figure 7 exhibits a comparison of the prediction performance of the proposed model and the other models on the forecasting data-set.

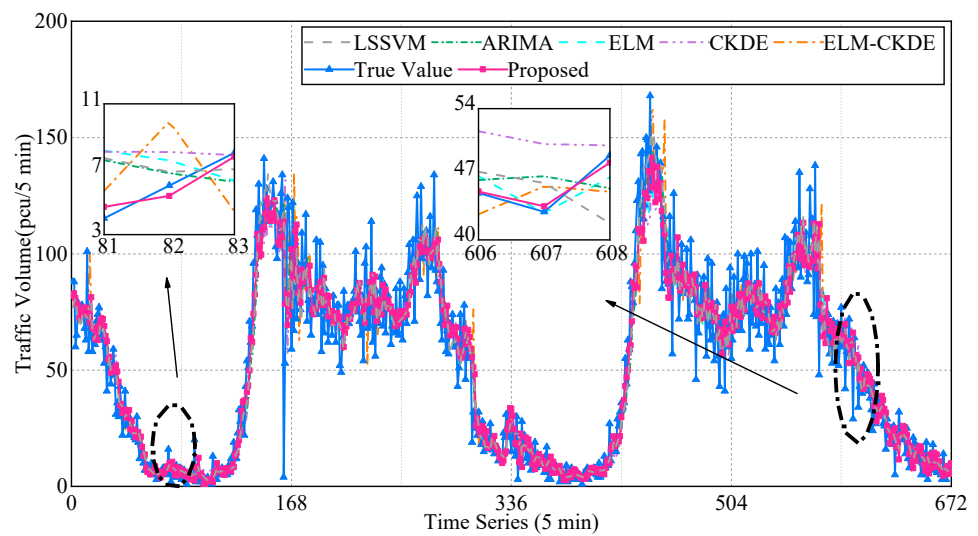


Figure 7. The prediction results of different models (dataset 1).

As shown in Figure 7, the predicted values of the proposed model are closer to the true value than other models in the local interval, which indicates the superiority of the proposed model.

3.6. Additional Case

To further verify the applicability of the proposed model, another set of data with different periods and fluctuations (dataset 2) was used to prove that the ELM-AKDE-CKDE model can provide superior short-term forecasts of traffic flows. Analogously, the statistical results of dataset 2 are depicted in Figure 8 and Table 4.

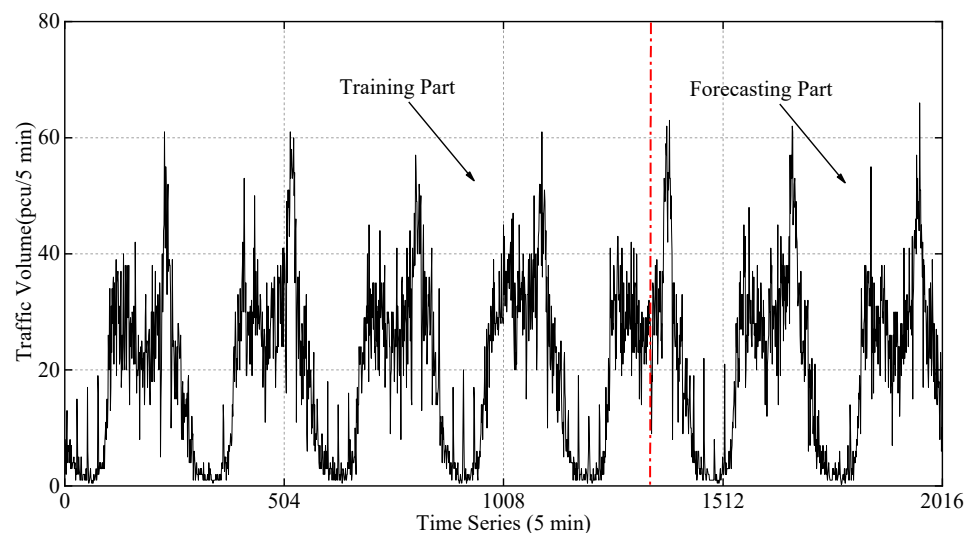


Figure 8. Traffic volume time series (dataset 2).

Table 4. The statistical information of the dataset 2.

Data Source	Mean	Std.	Maximum	Minimum	Skewness	Kurtosis
Dataset 2	20.1518	14.4026	66	0.25	−0.302	−0.693

In comparing dataset 1 and dataset 2, we can intuitively observe from Figures 4 and 8 that they have similar trends and show strong cyclical characteristics. Still, the average traffic flow of dataset 2 is slightly low, implying slightly less volatility. Analogously, the

same experiment was conducted on dataset 2, and the performance is shown in Table 5. In order to visualize the difference between the test performances of different methods, we constructed a bar graph according to Table 5, as shown in Figure 9. The intuitive comparative results are exhibited in Figure 10.

Table 5. Result comparisons of different models (dataset 2).

Model	MAE	MRPE	RMSE	RMSRE
Proposed	4.451	0.387	6.245	0.756%
ARIMA	4.554	0.394	6.391	0.748%
ELM-CKDE	4.595	0.417	6.396	0.826%
ELM	4.507	0.413	6.294	0.817%
CKDE	4.609	0.489	6.401	0.996%
LSSVM	4.492	0.397	6.314	0.781%

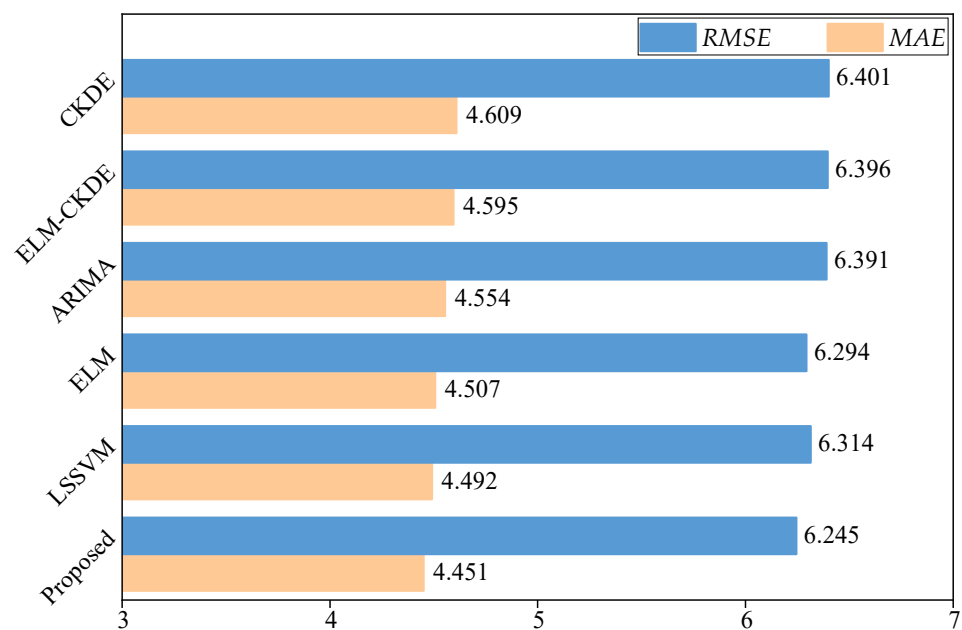


Figure 9. Performance comparison bar chart (dataset 2).

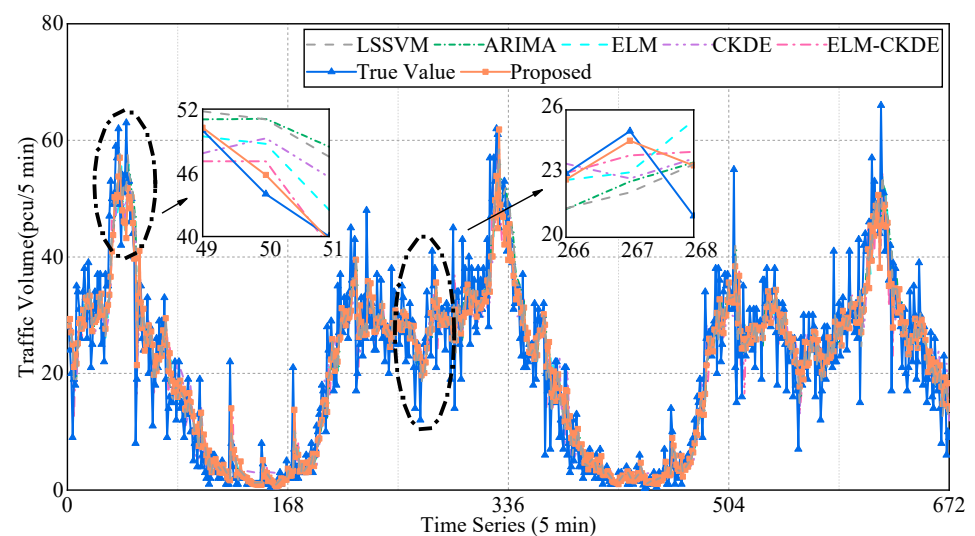


Figure 10. The prediction results of different models (dataset 2).

In Tables 2 and 5, it can be seen that the *MAE*, *RMSE*, and *RMSRE* of the proposed method of dataset 2 are smaller than the proposed method of dataset 1, but the *MRPE* is greater. This may have been due to the existence of different data characteristics in the two datasets, such as the smaller average traffic flow in dataset 2. It is worth noting that the *MRPE* of the ARIMA model is higher than that of the ELM-AKDE-CKDE model, but the *RMSRE* is lower than that of the proposed model. The reason could be attributed to the *RMSRE* indicator being more sensitive to outliers.

From the results presented in Table 5 and Figures 9 and 10, the conclusions are similar to the results of dataset 1. Namely, the proposed method outperformed the other five methods in the overall performance of the prediction task. Firstly, the AKDE-CKDE was better than CKDE in facilitating the prediction of stochastic traffic flow data. In addition, the hybrid model was superior to individual models because it could integrate the advantages of each model component. At the same time, the proposed model can well explain the nonlinear features and random features hidden in traffic flow data and has excellent adaptability.

To sum up, our ELM-AKDE-CKDE method performed the best for both datasets in terms of all metrics. This proves that the capabilities of the proposed method for modeling nonlinear and complex characteristic data are superior. The proposed model considered nonlinearity, nonstationarity, and randomness simultaneously, and thus achieved better prediction results than the single model that considered only linear or nonlinear. Our model thus further reduced the prediction errors and can be applied to predict short-term traffic flow accurately.

4. Conclusions

Since actual traffic flow sequences are affected by random factors, obtaining accurate traffic flow prediction results is often a significant challenge. In order to cope with these challenges, a novel hybrid prediction method based on ELM, AKDE, and CKDE was proposed and investigated in this study. It offers a way to improve the CKDE method by using the adaptive bandwidth method. To the best of our knowledge, the method was first applied to the field of short-term traffic flow prediction. The results prove that the AKDE-CKDE model has a more positive effect than CKDE in terms of improving prediction accuracy. Moreover, case studies based on the measured data illustrate that its performance was better than other models, including ARIMA, LSSVM, ELM-CKDE, ELM, and CKDE.

The novelty of this article is that the hybrid method can take into account nonlinear and stochastic characteristics embedded in traffic flow data and exhibit a satisfactory performance. Similar to other prediction methods, the proposed method also needs further improvement. It is worth noting that the method established in this paper does not decompose the traffic flow, and the short-term traffic flow prediction after decomposition is worth studying further. In addition, the characteristics of traffic flow data should be analyzed to provide a basis for selecting prediction models.

Author Contributions: Conceptualization, L.Z. and Y.B.; methodology, Y.B. and Y.W.; software, L.Z. and S.Z.; validation, L.Z., Y.B., Y.W. and S.Z.; formal analysis, investigation resources, and data curation, L.Z. and Y.B.; writing—original draft preparation, Y.B.; writing—review and editing, L.Z., Y.W. and S.Z.; visualization, Y.B., W.Z. and J.K.; supervision, W.Z.; project administration, J.K.; funding acquisition, L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Joint Training Base Construction Project for Graduate Students in Chongqing (No. JDLHPYJD2021016), the Group Building Scientific Innovation Project for Universities in Chongqing (No. CXQT21021), the Technology Research Project Fund of Chongqing Education Commission (No. KJQN202100712).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Han, L.; Huang, Y. Short-term traffic flow prediction of road network based on deep learning. *IET Intell. Transp. Syst.* **2020**, *14*, 495–503. [[CrossRef](#)]
2. Li, J.; Guo, F.; Sivakumar, A.; Dong, Y.; Krishnan, R. Transferability improvement in short-term traffic prediction using stacked LSTM network. *Transp. Res. Part C Emerg. Technol.* **2021**, *124*, 102977. [[CrossRef](#)]
3. Ma, T.; Antoniou, C.; Toledo, T. Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast. *Transp. Res. Part C Emerg. Technol.* **2020**, *111*, 352–372. [[CrossRef](#)]
4. Zhang, H.; Wang, X.; Cao, J.; Tang, M.; Guo, Y. A multivariate short-term traffic flow forecasting method based on wavelet analysis and seasonal time series. *Appl. Intell.* **2018**, *48*, 3827–3838. [[CrossRef](#)]
5. Luo, X.; Niu, L.; Zhang, S. An algorithm for traffic flow prediction based on improved SARIMA and GA. *KSCE J. Civ. Eng.* **2018**, *22*, 4107–4115. [[CrossRef](#)]
6. Shi, G.; Guo, J.; Huang, W.; Williams, B.M. Modeling seasonal heteroscedasticity in vehicular traffic condition series using a seasonal adjustment approach. *J. Transp. Eng.* **2014**, *140*, 1053–1058. [[CrossRef](#)]
7. Guo, J.; Huang, W.; Williams, B.M. Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. *Transp. Res. Part C Emerg. Technol.* **2014**, *43*, 50–64. [[CrossRef](#)]
8. Cai, L.; Zhang, Z.; Yang, J.; Yu, Y.; Zhou, T.; Qin, J. A noise-immune Kalman filter for short-term traffic flow forecasting. *Phys. Stat. Mech. Appl.* **2019**, *536*, 122601. [[CrossRef](#)]
9. Qi, Y.; Ishak, S. A Hidden Markov Model for short term prediction of traffic conditions on freeways. *Transp. Res. Part C Emerg. Technol.* **2014**, *43*, 95–111. [[CrossRef](#)]
10. Jiang, Y.; Huang, G.; Yang, Q.; Yan, Z.; Zhang, C. A novel probabilistic wind speed prediction approach using real time refined variational model decomposition and conditional kernel density estimation. *Energy Convers. Manag.* **2019**, *185*, 758–773. [[CrossRef](#)]
11. Wang, Y.; Zhao, L.; Li, S.; Wen, X.; Xiong, Y. Short term traffic flow prediction of urban road using time varying filtering based empirical mode decomposition. *Appl. Sci.* **2020**, *10*, 2038. [[CrossRef](#)]
12. Ryu, U.; Wang, J.; Kim, T.; Kwak, S.; Juhyok, U. Construction of traffic state vector using mutual information for short-term traffic flow prediction. *Transp. Res. Part C Emerg. Technol.* **2018**, *96*, 55–71. [[CrossRef](#)]
13. Jiang, Y.; Zhao, N.; Peng, L.; Xin, J.; Liu, S. Fast simulation of fully non-stationary wind fields using a new matrix factorization assisted interpolation method. *Mech. Syst. Signal Process.* **2022**, *172*, 108973. [[CrossRef](#)]
14. Chen, Q.; Song, Y.; Zhao, J. Short-term traffic flow prediction based on improved wavelet neural network. *Neural. Comput. Appl.* **2021**, *33*, 8181–8190. [[CrossRef](#)]
15. Yang, H.J.; Hu, X. Wavelet neural network with improved genetic algorithm for traffic flow time series prediction. *Optik* **2016**, *127*, 8103–8110. [[CrossRef](#)]
16. Wang, Q.M.; Fan, A.W.; Shi, H.S. Network traffic prediction based on improved support vector machine. *Int. J. Syst. Assur. Eng. Manag.* **2017**, *8*, 1976–1980. [[CrossRef](#)]
17. Wu, C.H.; Ho, J.M.; Lee, D.T. Travel time prediction with support vector regression. *IEEE Trans. Intell. Transp. Syst.* **2004**, *5*, 276–281. [[CrossRef](#)]
18. Ma, X.; Tao, Z.; Wang, Y.; Yu, H.; Wang, Y. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transp. Res. Part C Emerg. Technol.* **2015**, *54*, 187–197. [[CrossRef](#)]
19. Zheng, Z.; Chen, W.; Wu, X.; Chen, P.; Liu, J. LSTM network: A deep learning approach for short-term traffic forecast. *IET Intell. Transp. Syst.* **2017**, *11*, 68–75.
20. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [[CrossRef](#)]
21. Adnan, R.M.; Mostafa, R.R.; Kisi, O.; Yaseen, Z.M.; Shahid, S.; Zounemat-Kermani, M. Improving streamflow prediction using a new hybrid ELM model combined with hybrid particle swarm optimization and grey wolf optimization. *Knowl. Based. Syst.* **2021**, *230*, 107379. [[CrossRef](#)]
22. Cai, W.; Yang, J.; Yu, Y.; Qin, J. PSO-ELM: A hybrid learning model for short-term traffic flow forecasting. *IEEE Access* **2020**, *8*, 6505–6514. [[CrossRef](#)]
23. Cui, Z.; Huang, B.; Dou, H.; Tan, G.; Zheng, S.; Zhou, T. A hybrid learning model for short-term traffic flow forecasting. *IET Intell. Transp. Syst.* **2021**, *16*, 41–52. [[CrossRef](#)]
24. Diao, Z.; Zhang, D.; Wang, X.; Xie, K.; He, S.; Lu, X.; Li, W. A hybrid model for short-term traffic volume prediction in massive transportation systems. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 935–946. [[CrossRef](#)]
25. Guo, Z.; Zhao, X.; Chen, Y.; Wu, W.; Yang, J. Short-term passenger flow forecast of urban rail transit based on GPR and KRR. *IET Intell. Transp. Syst.* **2019**, *13*, 1374–1382. [[CrossRef](#)]
26. Zhou, B.; Ma, X.; Luo, Y.; Yang, D. Wind power prediction based on LSTM networks and nonparametric kernel density estimation. *IEEE Access* **2019**, *7*, 165279–165292. [[CrossRef](#)]
27. Jeon, J.; Taylor, J.W. Using conditional kernel density estimation for wind power density forecasting. *J. Am. Stat. Assoc.* **2012**, *107*, 66–79. [[CrossRef](#)]

28. Bessa, R.J.; Miranda, V.; Botterud, A.; Wang, J.; Constantinescu, E.M. Time adaptive conditional kernel density estimation for wind power forecasting. *IEEE Trans. Sustain. Energy* **2012**, *3*, 660–669. [[CrossRef](#)]
29. Botev, Z.I.; Grotowski, J.F.; Kroese, D.P. Kernel density estimation via diffusion. *Ann. Stat.* **2010**, *38*, 2916–2957. [[CrossRef](#)]
30. Cao, J.; Lin, Z.; Huang, G.B. Composite function wavelet neural networks with extreme learning machine. *Neurocomputing* **2010**, *73*, 1405–1416. [[CrossRef](#)]
31. Zambom, A.Z.; Dias, R. A Review of Kernel Density Estimation with Applications to Econometrics. *arXiv* **2012**, arXiv:1212.2812.
32. Zhao, L.; Wen, X.; Wang, Y.; Shao, Y. A novel hybrid model of ARIMA-MCC and CKDE-GARCH for urban short-term traffic flow prediction. *IET Intell. Transp. Syst.* **2022**, *16*, 206–217. [[CrossRef](#)]