Forecasting the Short-Term Metro Ridership With Seasonal and Trend Decomposition Using Loess and LSTM Neural Networks

DEWANG CHEN1,2, (Senior Member, IEEE), JIANHUA ZHANG1, AND SHIXIONG JIANG1,2
1College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China
2Key Laboratory of Intelligent Metro of Universities in Fujian Province, Fuzhou University, Fuzhou 350108, China
Corresponding author: Shixiong Jiang (sxjiang@fzu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61976055, and in part by the Foundation of Key Laboratory of Intelligent Metro of Universities in Fujian Province under Grant 53001703 and Grant 50013203.

ABSTRACT Forecasting the short-term metro ridership is an important issue for operation management of metro systems. However, it cannot be solved well by the single long short-term memory (LSTM) neural network alone for the irregular fluctuation caused by various factors. This paper proposes a hybrid algorithm (STL-LSTM) which combines the addition mode of Seasonal-Trend decomposition based on Loess (STL) and the LSTM neural network to mitigate the influences of irregular fluctuation and improve the performance of short-term metro ridership prediction. First, the original series is decomposed into three sub-series by the addition mode of STL. Then, the LSTM neural network is employed to predict each decomposed series. Finally, all the predicted outputs are merged as the overall output. The results show that the STL-LSTM model can achieve higher accuracy than the single LSTM model, support vector regression (SVR), and the EMD-LSTM model which combines the empirical mode decomposition and the LSTM neural network.

INDEX TERMS Short-term metro ridership prediction, seasonal-trend decomposition based on loess (STL), long short-term memory (LSTM) neural network.

I. INTRODUCTION
The popularity of private cars has led to increasingly congested roads and an increase in CO2 emissions. From an ecological perspective, it would be better to use public transport rather than private cars. The government increases investment in constructing metro for its unique advantages. First, a well-designed metro system can save travel cost and improve reliability by reducing traffic congestion on the roads. Second, the railway is a safe form as the chances of accidents and breakdown of railways are minimum as compared to other modes of transport. Besides, traveling by metro can save energy and reduce air pollution. It is doubtless that increasing investment in metro can improve traffic condition. The quality of metro services has been concerned by more and more passengers with the rapid development of metro system. As the first choice of mass passenger, the issues including insufficient capacity and trampling must be solved urgently. The ridership is the fundamental basis of metro services. Accurate metro ridership prediction can get the real-time situation of ridership and reflect the actual travel demand. It provides an important basis for traffic regulation and emergency warning. Besides, it can be used to adjust metro planning and minimize operating costs.

Currently, theoretical studies on ridership forecasting algorithms mainly include: neural networks [1], [2], support vector machine [3], multiscale radial basis function [4], autoregressive integrated moving average [5], nonlinear autoregressive with external input [6], distance-decay weighted regression [7], K-nearest neighbour pattern matching [8], [9], network Kriging method [10], gradient boosting decision trees [11], hybrid loss function [12] and so on. But the prediction accuracy of a single model on some datasets is limited. We need to make some attempts to improve prediction accuracy. The combination of multiple models may have a good result [13]–[15]. For time series data, it is promising to combine decomposition model and prediction model.

The decomposition of ridership data can eliminate the influence of periodic factors and reflect the real laws and trends of metro ridership, which can improve the accuracy of metro ridership prediction. Thus, some researchers tried...
to propose the hybrid model which combines decomposition method and prediction algorithm, such as a hybrid model combining ensemble empirical mode decomposition and gray support vector machine [16], a prediction model based on wavelet decomposition with support vector machine [17], an empirical mode decomposition based long short-term memory neural network forecasting model [18], a hybrid model based on ARIMA and wavelet decomposition [19], a new model based on long short-term memory network and Gaussian mixture model [20], a novel model based on variational mode decomposition and other outstanding models [21], [22].

Metro ridership might be influenced by holidays and major events easily. Taking the annual average ridership as a baseline, daily ridership on the Spring Festival is lower than it, but the one on the National Day is higher than it. The celebrations attract a great number of people which leads to an increase in metro ridership. While the metro ridership will drop sharply if there are cases of extreme weather. Besides, it is obvious that metro ridership owns a typical weekly pattern. Those irregular fluctuations bring the single model some troubles. The above decomposition models can decompose series and improve prediction accuracy, but they do not make full use of the characteristics of time series if applied in the field of metro ridership prediction.

The seasonal-trend decomposition based on loess (STL) fully considers the characteristics of ridership data and decomposes it to seasonal, trend and residual components. It can perform well in prediction [23], [24] and could be used in exponential smoothing [25]. Besides, it was used in long-term traffic prediction recently [26], but the authors did not discuss the impact of frequency of decomposition and the time-step of LSTM neural network.

In this paper, the STL is used to separate the seasonal series, trend series and residual series from metro ridership data. The LSTM neural network is employed to forecast three decomposed series respectively. Finally, the predicted values of each LSTM model are added up to attain the predicted ridership. In addition, the prediction results of single STL, STL-LSTM, EMD-LSTM and SVR are compared.

The remainder of this paper is structured as follows. Section II introduces the STL, LSTM and SVR models. In Section III, the hybrid algorithm of the STL-LSTM, EMD-LSTM and SVR are described. A case analysis is proposed, and the experimental results are provided in Section IV. Finally, the conclusions are presented and limitations of this algorithm are discussed.

II. METHODOLOGY
A. SEASONAL-TREND DECOMPOSITION BASED ON LOESS
STL is a time series decomposition method which is based on locally weighted scatterplot smoothing (loess) [27]. \( \hat{g}(x) \) is a loess regression curve which is used for smoothing. It can be computed in the following way. Suppose \( x_i \) and \( y_i \) for \( i = 1 \) to \( n \) are measurement of a dependent variable, respectively.

At first, choose an integer \( q \) as the number of values that are closest to \( x \). And then set a neighborhood weight for \( x_i \) according to the distance between \( x_i \) and \( x \) when \( q \leq n \). \( W \) denotes a weight function.

\[
W(u) = \begin{cases} 
(1 - u^3)^3 & \text{for } 0 \leq u < 1 \\
0 & \text{for } u \geq 1
\end{cases}
\]  

(1)

\( v_i(x) \) is the neighborhood weight of \( x_i \), \( \lambda_q(x) \) is the distance between \( x_i \) and \( x \):  

\[
v_i(x) = W(|x_i - x|) / \lambda_q(x)
\]  

(2)

The longer the distance \( \lambda_q(x) \) is, the smaller the weight value \( v_i(x) \) is. Finally, find a fitted value \( \hat{g}(x) \) at point \( x \) with weight \( v_i(x) \).

STL is made up of two procedures, including an inner loop and an outer loop. The inner loop is nested inside the outer loop. The main steps of inner loop are seasonal smoothing and trend smoothing. The detailed steps of \( k \)th epoch are as follows:

Step 1: Detrending. Get a new series by subtracting trend values \( T_k \) from original values \( Y \).

Step 2: Cycle-subseries Smoothing. Each cycle-subserie obtained from Step 1 is regressed by loess, the result is recorded as \( C_{k+1} \).

Step 3: Low-Pass Filtering. The filter for \( C^k_{k+1} \) includes three steps. The first step is a moving average of length \( n \). \( n \) is the number of samples. The followed step is also a moving average of length \( n \). The last step is a moving average of length \( 3 \). And then, the loess is applied to the results of low-pass filtering. The result is recorded as \( L^k_{k+1} \).

Step 4: Detrending. Get seasonal series \( S^{k+1}_v \) from \( C_{k+1} \) minus \( L^k_{k+1} \).

\[
S^{(k+1)}_v = C^{(k+1)}_v - L^{(k+1)}_v \quad \text{for } v = 1 \text{ to } n
\]  

(3)

Step 5: Deseasonalizing. Get deseasonalized series from \( Y \) minus \( S \).

Step 6: Trend Smoothing. The trend series \( T^{(k+1)}_v \) is got after the loess is applied to deseasonalized series.

The steps of outer loop are as follows:

The values of \( T_v \) and \( S_v \) are got after the inner loop. Then the residual series \( R_v \) can be calculated according to Eq. 4:

\[
R_v = Y_v - T_v - S_v
\]  

(4)

The robustness weight \( \rho \) is defined for evaluating the robustness of \( R_v \). \( \rho_v \) is the robustness weight at time \( v \).

\[
h = 6 \text{median}(|R_v|)
\]  

(5)

\[
\rho_v = B(|R_v|)/h
\]  

(6)

The formula of bisquare weight function \( B \) is as follows:

\[
B(u) = \begin{cases} 
(1 - u^2)^2 & \text{for } 0 \leq u < 1 \\
0 & \text{for } u \geq 1
\end{cases}
\]  

(7)
B. LONG SHORT-TERM MEMORY NEURAL NETWORK

The traditional neural network has been plagued by the inability to interpret input sequences that depend on information and context. The Recurrent Neural Network (RNN) feeds back the output to the input to provide the context of the last input to solve this problem [28]. However, RNN has a considerable disadvantage, which will lead to the problem of vanishing gradient. The LSTM neural network is an improved algorithm whose neurons can keep memory in their channels to mitigate the vanishing gradient problem effectively [29]. That makes LSTM neural network become popular in time series prediction [30]–[32].

The key to LSTM neural network is the cell state. The LSTM can remove or add information to the cell state through the three gates.

1) FORGET GATE

The forget gate can decide how much information obtained from the previous moment can be retained at current moment.

\[ f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \]  

where \( W_f \) is the weight matrix of the forgetting gate, \( x_t \) is the current input, \( h_{t-1} \) is the previous output of memory block, \( b_f \) is the bias term of the forgetting gate, and \( \sigma \) is the sigmoid function.

2) INPUT GATE

The input gate decides how much information obtained from the current input \( x_t \) can be saved in the cell state \( c_t \):

\[ i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \]  

3) OUTPUT GATE

The output gate is closely related to the output value \( h_t \).

\[ o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \]  

The entire transition process of old cell \( c_{t-1} \) to the new state \( c_t \) is as follows. The tanh is an active function.

\[ \tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \]  

\[ c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \]  

\[ h_t = o_t \cdot \tanh(c_t) \]  

The structure of three gates is described as Fig. 1.

C. SUPPORT VECTOR REGRESSION

Support vector machine (SVM) is proposed for binary classification, and SVR is an important application branch of SVM.

The regression expression is as follows.

\[ f(x) = w^T \Phi(x) + b \]  

SVR can be formalized as follows.

\[ \min_{w,b} ||w||^2 / 2 + C \sum_{i=1}^{n} l_{e}(f(x_i), y_i) \]  

where \( C \) is a constant value for regularization, and \( l_{e} \) is a function of \( \varepsilon \)-insensitive loss. The loss is only calculated when the difference between \( f(x_i) \) and \( y_i \) is greater than \( \varepsilon \).

\[ l_{e}(z) = \begin{cases} 0 & \text{if } |z| < \varepsilon \\ |z| - \varepsilon & \text{otherwise} \end{cases} \]

The utilized function is radial basis function.

III. PROCEDURAL FRAMEWORK

A. STL-LSTM PREDICTION MODEL

The STL-LSTM is a hybrid algorithm which combines the addition mode of STL and LSTM neural network to improve prediction performance.

The specific steps are as follows:

Step 1: Calculate the number of metro station ridership during the statistical period to obtain the original series.

Step 2: Set the decomposition period and use the STL to decompose the original series into three sub-series.

Step 3: The data is divided into training set and test set, and the time-step and predict-step are set before starting training. The three sub-series obtained by Step 2 are trained using the same parameters of the LSTM neural network, and their test sets are predicted separately.

Step 4: Add the prediction results of the sub-series as the prediction results of the original data, and then calculate the \( R^2 \), mean absolute error (MAE), and root mean square error (RMSE).

\[ R^2 = 1 - \frac{\sum_{i=0}^{n} (Q'_i - Q_i)^2}{\sum_{i=0}^{n} (\hat{Q}_i - Q_i)^2} \]  

\[ \text{MAE} = \frac{\sum_{i=0}^{n} |(Q'_i - Q_i)|}{n} \]  

\[ \text{RMSE} = \sqrt{\frac{\sum_{i=0}^{n} (Q'_i - Q_i)^2}{n}} \]

where \( Q' \) denotes the predicted value, \( Q \) is the real value, \( \hat{Q} \) is the mean value of \( Q \), and \( n \) is the number of data in the test set.

Step 5: Adjust parameters to find the optimal frequency of decomposition and the time-step of LSTM until indexes cannot attain a massive improvement.
The flow chart of STL-LSTM is presented in Fig. 2.

**FIGURE 2. Flow chart of STL-LSTM.**

**B. EMD-LSTM PREDICTION MODEL**

EMD-LSTM is a similar model which combines the EMD and the LSTM neural networks for short-term load forecasting, financial time series forecasting, foreign exchange rates forecasting [33]–[35] and so on.

First, the original series is decomposed into several sub-series by the EMD. Second, establish the LSTM prediction models for the sub-series to get the prediction value respectively. Finally, reconstruct the prediction values and get the final result.

**IV. CASE STUDY**

**A. STATISTICS**

The original series is the daily metro ridership of Fuzhou Metro Line 1 in 2018 calculated from the AFC transaction data. The average metro ridership in 2018 is 167,000. During holidays, the number of metro ridership was lower than the annual average one during the Spring Festival, Qingming Festival, Dragon Boat Festival and Mid-Autumn Festival but higher during the other holidays. Especially during the Spring Festival, many metro passengers have not returned to the city yet, and the cold and rainy weather during holidays is not conducive to public travel. Besides, the influence of typhoon makes the ridership on July 11 become the lowest. The celebration of the Lantern Festival held at Dongjiekou attracted a large number of passengers, leading to the highest metro ridership on March 2, 2018. These reasons lead to many irregular fluctuations in the dataset.

Fig. 3 shows the original series utilised in this paper.

The periodicity of metro ridership is also obvious, especially the values of autocorrelation function (ACF) and partial autocorrelation function (PCAF) are rising sharply when the Lag at multiples of seven as shown in Fig. 4.

**FIGURE 3. The daily metro ridership of Fuzhou Metro in 2018.**

**FIGURE 4. ACF and PCAF.**

**B. PROBLEM STATEMENT**

This study tries to establish a model which can mitigate the influences of irregular fluctuation and improve the performance of short-term metro ridership prediction.

In the following section, first, we decompose the original series via STL. Second, a single LSTM prediction model, an SVR model, an STL-LSTM hybrid prediction model and an EMD-LSTM model are used to predict the ridership in the test set. Finally, the experimental results will be analysed to select the better model.

**C. THE FREQUENCY OF DECOMPOSITION**

The frequency of decomposition has an impact in the final results. Fig. 5 shows the decomposition results of STL.

The mean relative error (MRE) is used to select the frequency of decomposition.

\[
MRE = \frac{1}{n} \sum_{i=0}^{n} \left| \frac{Q_i' - Q_i}{Q_i} \right| \quad (20)
\]

where \( Q_i' \) denotes the predicted value, \( Q_i \) is the real value and \( n \) is the number of data in the test set.

It’s found that the improvement is not obvious after the frequency of decomposition is more than 10 from Fig. 6. So this paper just discusses the results of frequency range during 2-10.

**D. THE DECOMPOSITION RESULT OF EMD**

The difference between EMD-LSTM and STL-LSTM is the decomposition method. Fig. 7 shows the data distribution after the original series is decomposed by the EMD.
Six intrinsic mode functions (IMFs) and one residue are obtained after the original series decomposed by EMD.

E. PARAMETERS OF LSTM
All experiments involving LSTM use the following parameters. The training set consists of 70% of data and the test set is the remaining 30%. The number of hidden layers of LSTM is 1. There are four neurons in the hidden layer. The activation function is tanh. The loss function is mean squared error. The optimizer is adaptive moment estimation. The number of neurons in the output layer is 1.

F. PREDICTION RESULTS
The highest accuracy occurs when the frequency of decomposition is set as 2 (see Fig. 6), so the index of STL-LSTM in this section is all got when the frequency of decomposition is 2.

| Time-step | LSTM | STL-LSTM | EMD-LSTM | SVR |
|-----------|------|----------|----------|-----|
| 1         | 9117.1561 | 7636.9281 | 9524.8664 | 9907.2058 |
| 2         | 9100.8738 | 6620.8328 | 9443.3128 | 9345.8429 |
| 3         | 10390.205 | 5585.7636 | 9216.3456 | 9466.6717 |
| 4         | 9543.4333 | 4137.2949 | 8786.1940 | 9517.9731 |
| 5         | 11141.2056 | 5138.6389 | 9347.1771 | 9564.2684 |
| 6         | 11278.204 | 4948.5718 | 9831.0393 | 9076.3920 |
| 7         | 9266.2956 | 4581.1316 | 9069.7149 | 7531.5321 |
| 8         | 9540.9256 | 6126.9756 | 9173.5594 | 7471.1646 |
| 9         | 8652.2669 | 4947.3217 | 9502.8573 | 7455.8217 |
| 10        | 8659.2826 | 4404.2034 | 8978.5526 | 7502.7299 |
| 11        | 8486.8130 | 4135.8790 | 8924.7132 | 7452.5236 |
| 12        | 8817.6044 | 2802.3488 | 8526.0358 | 7494.5554 |
| 13        | 8775.9799 | 3678.1556 | 8599.1762 | 7332.3960 |
| 14        | 8499.4386 | 4691.0034 | 8381.9518 | 7257.1146 |
| 15        | 8494.6084 | 3583.4950 | 8836.7332 | 7207.3648 |
| 16        | 8419.6223 | 3248.8404 | 9078.2975 | 7592.7522 |
| 17        | 8767.2510 | 3589.3040 | 9870.7802 | 7489.2235 |
| 18        | 8180.2257 | 3595.7105 | 8489.9693 | 7510.1964 |
| 19        | 7872.2488 | 3329.3015 | 8686.4264 | 7830.9757 |
| 20        | 8115.1451 | 4385.9975 | 9331.6952 | 7672.4110 |
| 21        | 7833.3651 | 3590.7659 | 7878.3969 | 7442.0125 |
| 22        | 7731.0694 | 3107.5385 | 9481.3590 | 7936.7537 |
| 23        | 8681.1893 | 3376.0400 | 7549.8721 | 7609.8580 |
| 24        | 8076.6169 | 3777.3732 | 7638.2356 | 7536.1844 |
| 25        | 9371.9050 | 3609.2496 | 8314.6073 | 7687.0943 |
| 26        | 8128.6994 | 3067.2324 | 7418.1185 | 7729.9318 |
| 27        | 7783.1036 | 3334.9817 | 8650.9903 | 7734.5517 |
| 28        | 8272.8848 | 6235.5186 | 7776.2768 | 7792.3520 |
| 29        | 8226.0184 | 3521.1644 | 7208.2024 | 7832.4447 |
| 30        | 11298.4144 | 6259.8363 | 8130.5658 | 7734.1467 |
| 31        | 8893.2821 | 3184.1912 | 7333.5667 | 7610.1895 |
| 32        | 10084.9708 | 3275.3645 | 7123.4976 | 7630.2147 |
| 33        | 7842.6256 | 5636.5259 | 7634.7227 | 7844.9106 |
| 34        | 7651.8581 | 3967.9417 | 7376.3353 | 7856.8723 |
| 35        | 9496.8533 | 4785.2877 | 7869.6593 | 7751.6579 |
| 36        | 8475.2645 | 5092.0359 | 7364.2314 | 7868.2399 |
| 37        | 7819.8649 | 3069.3705 | 7386.1992 | 7864.2016 |
| 38        | 10131.8065 | 3328.0669 | 7763.9184 | 8139.4138 |
| 39        | 7377.7005 | 3353.7187 | 7648.8219 | 8223.7148 |
| 40        | 10610.124 | 5571.1826 | 9026.0607 | 8341.0228 |
| 41        | 8263.1736 | 4070.2213 | 7422.3739 | 8896.3128 |
| 42        | 11764.6128 | 4897.4243 | 7734.2527 | 10233.9008 |
| 43        | 8429.1312 | 3612.8965 | 8145.2430 | 10008.0006 |
| 44        | 8657.3284 | 5190.7270 | 7666.1757 | 9411.3475 |
| 45        | 7646.8204 | 3477.9917 | 7716.4278 | 8828.3688 |
| 46        | 8273.1395 | 3370.3366 | 7958.8673 | 8885.2492 |
| 47        | 9833.5782 | 3586.7283 | 7640.2888 | 8967.7112 |
| 48        | 10236.3919 | 4504.7185 | 8341.5160 | 9250.9964 |
| 49        | 9552.1996 | 2632.2982 | 7142.3789 | 9416.9578 |
| 50        | 9268.9912 | 4855.9714 | 7172.9381 | 9503.6957 |
| mean      | 8944.7122 | 4282.6047 | 8307.3666 | 8204.7705 |
We choose $R^2$ to compare the fitting precision of different model. The closer the value is to 1, the better the performance. It can be found from Fig. 8 that the single LSTM for this dataset does not perform well, the SVR can perform better, but the SVR underperform EMD-LSTM, and STL-LSTM can get highest rating at every time-step. As shown in Table 1, the MAE of the STL-LSTM is the lowest in those models. The mean value of STL-LSTM is a little smaller than half of the value of LSTM. But there is little influence on the final single-step prediction accuracy.

As shown in Table 1, the MAE of the STL-LSTM is the lowest in those models. The mean value of STL-LSTM is a little smaller than half of the value of LSTM. But there is little improvement in EMD-LSTM and SVR.

We can find that the RMSE of the STL-LSTM is also the lowest in Table 2. There is an interesting fact that the MAE of SVR is smaller than that of EMD-LSTM, but the RMSE of SVR is larger than that of EMD-LSTM, that means the deviations of some prediction values in SVR are much larger than those in EMD-LSTM.

V. CONCLUSION

In this paper, a hybrid algorithm that combines the addition mode of STL and the LSTM neural network was proposed to improve the performance of short-term metro ridership prediction. The STL decomposed the ridership data into seasonal, trend and residual series. And then the LSTM prediction model was established for each of the three sub-series. Finally, the accurate prediction of metro ridership was achieved by adding up all the prediction results.

The study found that the frequency of decomposition has an influence on the final single-step prediction accuracy. Generally speaking, smaller frequency of decomposition can achieve higher prediction accuracy. In addition, the indexes of STL-LSTM are much better than other models, which denotes the effectiveness of decomposition. Compared with EMD-LSTM, the prediction accuracy of STL-LSTM is also improved in the test set, which means that STL-LSTM is more suitable for the decomposition of metropolitan transit authority.

In the future, we will try to use different data sources, set different parameters of LSTM and use different models for each decomposed series to explore the generalization and optimization of the algorithm. In addition, it is a good direction to apply optimization methods for adaptive step size in the model to improve the forecasting accuracy further.

REFERENCES

[1] L. Liu and R.-C. Chen, “A novel passenger flow prediction model using deep learning methods,” Transp. Res. C, Emerg. Technol., vol. 84, pp. 74–91, Nov. 2017.

[2] Y. Liu, Z. Liu, and R. Jia, “DeepPF: A deep learning based architecture for urban transit authority,” Transp. Res. C, Emerg. Technol., vol. 101, pp. 18–34, Apr. 2019.

[3] S. Liu and E. Yao, “Holiday passenger flow forecasting based on the modified least-square support vector machine for the metro system,” J. Transp. Eng., A, Syst., vol. 142, no. 6, Jun. 2016, Art. no. 04016005.

[4] Y. Li, W. Song, S. Sun, X. Ma, and G. Lu, “Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks,” Transp. Res. C, Emerg. Technol., vol. 77, pp. 306–328, Apr. 2017.

[5] Z. Guan, J. Yang, and J. Li, “Forecast of short-term passenger flow of urban railway stations based on seasonal ARIMA model,” in Proc. 3rd Int. Conf. Elect. Inf. Technol. Rail Transp. (EITRT), vol. 483, 2018, pp. 759–767.

[6] C. Cui, H. Jia, L. Huang, and X. Zhang, “Fuzzy multivariate NARX model for passenger entrance flow prediction in the Shanghai subway system,” J. Transp. Eng., A, Syst., vol. 143, no. 2, Feb. 2017, Art. no. 04016005.

[7] J. Gutierrez, O. D. Cardozo, and J. C. Garcia-Palomares, “Transit ridership forecasting at station level: An approach based on distance-decay weighted regression,” J. Transp. Geography, vol. 19, no. 6, pp. 1081–1092, Nov. 2011.

[8] B. Yu, X. Song, F. Guan, Z. Yang, and B. Yao, “K-nearest neighbor pattern matching,” J. Transp. Geography, vol. 77, pp. 306–328, Apr. 2017.

[9] D. Zhang and X. Wang, “Transit ridership estimation with network kriging: A case study of second avenue subway, NYC,” J. Transp. Geogr., vol. 23, pp. 134–140, Apr. 2011.

[10] D. Zhang and X. Wang, “Transit ridership estimation with network kriging: A case study of second avenue subway, NYC,” J. Transp. Geogr., vol. 23, pp. 134–140, Apr. 2011.

[11] Z. Guang, J. Yang, and J. Li, “Forecast of short-term passenger flow of urban railway stations based on seasonal ARIMA model,” in Proc. 3rd Int. Conf. Elect. Inf. Technol. Rail Transp. (EITRT), vol. 483, 2018, pp. 759–767.

[12] C. Cui, H. Jia, L. Huang, and X. Zhang, “Fuzzy multivariate NARX model for passenger entrance flow prediction in the Shanghai subway system,” J. Transp. Eng., A, Syst., vol. 143, no. 2, Feb. 2017, Art. no. 04016005.

[13] Z. Guang, J. Yang, and J. Li, “Forecast of short-term passenger flow of urban railway stations based on seasonal ARIMA model,” in Proc. 3rd Int. Conf. Elect. Inf. Technol. Rail Transp. (EITRT), vol. 483, 2018, pp. 759–767.

[14] C. Cui, H. Jia, L. Huang, and X. Zhang, “Fuzzy multivariate NARX model for passenger entrance flow prediction in the Shanghai subway system,” J. Transp. Eng., A, Syst., vol. 143, no. 2, Feb. 2017, Art. no. 04016005.
[14] C. Ding, J. Duan, Y. Zhang, X. Wu, and G. Yu, “Using an ARIMA-GARCH modeling approach to improve subway short-term ridership forecasting accounting for dynamic volatility,” IEEE Trans. Intell. Transp. Syst., vol. 19, no. 4, pp. 1054–1064, Apr. 2018.

[15] X. Ma, J. Zhang, B. Du, C. Ding, and L. Sun, “Parallel architecture of convolutional bi-directional LSTM neural networks for network-wide metro ridership prediction,” IEEE Trans. Intell. Transp. Syst., vol. 20, no. 6, pp. 2278–2288, Jun. 2019.

[16] X. Jiang, L. Zhang, and X. M. Chen, “Short-term forecasting of high-speed rail demand: A hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in China,” Transp. Res. C, Emerg. Technol., vol. 44, pp. 110–127, Jul. 2014.

[17] Y. Sun, B. Leng, and W. Guan, “A novel wavelet-SVM short-time passenger flow prediction in beijing subway system,” Neurocomputing, vol. 166, pp. 109–121, Oct. 2015.

[18] Q. Chen, D. Wen, X. Li, D. Chen, H. Lv, J. Zhang, and P. Gao, “Empirical mode decomposition based short-term memory neural network forecasting model for the short-term metro passenger flow,” PloS ONE, vol. 14, no. 9, 2019, Art. no. e0222365.

[19] X. Wang, N. Zhang, Y. Chen, and Y. Zhang, “Short-term forecasting of urban rail transit ridership based on ARIMA and wavelet decomposition,” in Proc. AIP Conf., 2018, Art. no. 040025.

[20] J. Zhang, J. Yan, D. Infeld, Y. Liu, and F.-S. Lien, “Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model,” Appl. Energy, vol. 241, pp. 229–244, May 2019.

[21] Y. Zhang, S. Gao, J. Han, and M. Ban, “Wind speed prediction research considering wind speed ramp and residual distribution,” IEEE Access, vol. 7, pp. 131873–131887, 2019.

[22] Y. Zhang, Y. Zhao, C. Kong, and B. Chen, “A new prediction method based on VMD-PRBF-ARMA-E model considering wind speed characteristic,” Energy Convers. Manage., vol. 203, Jan. 2020, Art. no. 112254.

[23] M. Shamsudduha, R. E. Chandler, R. G. Taylor, and K. M. Ahmed, “Recent trends in groundwater levels in a highly seasonal hydrological system: The Ganges-Brahmaputra-Meghna delta,” Hydrof. Earth Syst. Sci., vol. 13, no. 12, pp. 2373–2385, 2009.

[24] M. Theodosiou, “Forecasting monthly and quarterly time series using STIL decomposition,” Int. J. Forecasting, vol. 27, no. 4, pp. 1178–1195, Oct. 2011.

[25] C. Bergmeir, R. J. Hyndman, and J. M. Benítez, “Bagging exponential smoothing methods using STIL decomposition and Box–Cox transformation,” Int. J. Forecasting, vol. 32, no. 2, pp. 303–312, Apr. 2016.

[26] Y. Huo, Y. Yan, D. Du, Z. Wang, Y. Zhang, and Y. Yang, “Long-term span traffic prediction model based on STIL decomposition and LSTM,” in Proc. 20th Asia-Pacific Netw. Oper. Manage. Symp. (APNOMS), Sep. 2019, pp. 1–4.

[27] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, “STL: A seasonal-trend decomposition,” J. Off. Statist., vol. 6, no. 1, pp. 3–73, 1990.

[28] R. H. Austin and J. J. Hopfield, “Attempted time-resolved measurement of photo-assisted charge transfer of cytochrome c-iron hexacyanide,” in Electron Transport and Oxygen Utilization, London, U.K.: Macmillan Education, 1982, pp. 73–80.

[29] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[30] R. Fu, Z. Zhang, and L. Li, “Using LSTM and GRU neural network methods for traffic flow prediction,” in Proc. 31st Youth Acad. Annu. Conf. Chin. Assoc. Autom. (YAC), Nov. 2016, pp. 324–328.

[31] T. Zhang, S. Song, S. Li, L. Ma, S. Pan, and L. Han, “Research on gas concentration prediction models based on LSTM multidimensional time series,” Energies, vol. 12, no. 1, p. 161, 2019.

[32] Y. Li, Z. Zhu, D. Kong, H. Han, and Y. Zhao, “EA-LSTM: Evolutionary attention-based LSTM for time series prediction,” Knowl.-Based Syst., vol. 181, Oct. 2019, Art. no. 104785.

[33] B. Zhang, “Foreign exchange rates forecasting with an EMD-LSTM neural networks model,” J. Phys., Conf. Ser., vol. 1053, Jul. 2018, Art. no. 012005.

[34] J. Cao, Z. Li, and J. Li, “Financial time series forecasting model based on CEEMDAN and LSTM,” Phys. A, Stat. Mech. Appl., vol. 519, pp. 127–139, Apr. 2019.

[35] L. Chen, Y. Chi, Y. Guan, and J. Fan, “A hybrid attention-based EMD-LSTM model for financial time series prediction,” in Proc. 2nd Int. Conf. Artif. Intell. Big Data (ICAIBD), May 2019, pp. 113–118.

**DEWANG CHEN** (Senior Member, IEEE) received the Ph.D. degree in control theory from the Chinese Academy of Sciences, Beijing, China, in 2003. In 2009, he was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, University of California at Berkeley, Berkeley, CA, USA. From 2011 to 2014, he was a Professor with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing. In 2012, he was a Visiting Researcher with the Center for Industrial and Applied Mathematics, University of South Australia, Adelaide, SA, Australia. He is currently a Minjiang Distinguished Professor with the College of Mathematics and Computer Science and the Founding Director of the Key Laboratory of Intelligent Metro of Universities in Fujian Province. He is the author of more than 100 articles and four monographs. His current research interests include AI, machine learning, big data, and intelligent transportation systems. Dr. Chen has served on program committees for more than 15 international conferences. He has been serving as a Senior Member and an Associate Editor for the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, since 2014.

**JIANHUA ZHANG** received the bachelor’s degree from Fuzhou University, Fuzhou, China, in 2018, where he is currently pursuing the master’s degree with the College of Mathematics and Computer Science. His research interests include machine learning and data mining in railway systems.

**SHIXIONG JIANG** received the B.S. degree in traffic and transportation and the Ph.D. degree in control science and engineering from Beijing Jiaotong University, in 2014 and 2019, respectively. He is currently a Lecturer with the College of Mathematics and Computer Science, Fuzhou University. His current research interests include public transportation planning and operation, data mining, and intelligent transportation systems.