Short-Term Passenger Flow Prediction With Decomposition in Urban Railway Systems

YANGYANG ZHAO\(^1\), ZHENLIANG MA\(^2\), YI YANG\(^1\), WENHUA JIANG\(^2\), AND XINGUO JIANG\(^1\)

\(^1\)Department of the Transportation and Logistics, Southwest Jiaotong University, Chengdu 610031, China
\(^2\)Institute of Transport Studies, Monash University, Clayton, VIC 3800, Australia

Corresponding author: Xinguo Jiang (xjiang@swjtu.edu.cn)

This work was supported in part by the Natural Science Foundation of China under Grant 71771191.

ABSTRACT

Accurate prediction of short-term passenger flow is vital for real-time operations control and management. Identifying passenger demand patterns and selecting appropriate methods are promising to improve prediction accuracy. This paper proposes a hybrid prediction model with time series decomposition and explores its performance for different types of passenger flows with varied characteristics in urban railway systems. The Seasonal and Trend decomposition using Loess (STL) is used to decompose the passenger flow into the seasonal, trend, and residual time series, representing the constant, long-term fluctuant, and stochastic passenger demand patterns. The approximate entropy (ApEn) is utilized to quantify the predictability of each component. After that, the Holt-Winters (HW) method is developed to predict the seasonal and trend components that both with high predictability. The long short-term memory networks (LSTM NN) are proposed to predict the residual with low predictability. The outputs from these models are combined to predict the short-term passenger flow. We assess the model performance with entry, exit, and transfer passenger flows at the target stations, which are selected via the agglomerative hierarchical clustering (AHC) algorithm in the Shanghai Metro. Compared with three representative models, the results show that 1) the decomposition-based hybrid model performs well for both one-step and multi-step predictions in terms of accuracy and robustness; 2) ApEn is useful for choosing appropriate prediction models.

INDEX TERMS

Short-term metro passenger flow prediction, seasonal and trend decomposition using Loess (STL)-based model, approximate entropy (ApEn), agglomerative hierarchical clustering (AHC) algorithm.

I. INTRODUCTION

Short-term demand prediction facilitates advanced online applications in public transport, including predictive operations control, service, and customized information provision to passengers [1]. Numerous short-term prediction models have been proposed in transportation for different applications, such as travel demand [2], traffic flow [3], travel time [4], and crowding [5]. Regardless of the varied application scenarios, the short-term prediction in transportation is essentially a time series forecasting problem with exogenous explanatory variables (e.g., weather, land use). In this context, the effectiveness of predictions depends on the match of time series patterns and model assumptions (capability in capturing patterns). The match (i.e., prediction performance) can be improved from perspectives of inputs (time series or variables transformation), functions (statistic models and machine learning models) or outputs (correcting or combining results). Depends on the features of the studied data, different prediction methods may need to be developed for the best performance. The paper focuses on the station-based short-term passenger flow prediction in urban railway systems.

From prediction inputs perspectives, many studies developed novel methods by adding, constructing, or transforming input variables or time series data of demand. For example, Pereira et al. [6] combined the contextual information about special events with entry/exit flow data to predict public transport arrivals; Ni et al. [7] adopted the rates of social media posts and passenger flow information to estimate passenger demand. Several studies developed novel statistical models to construct or transform the time series of demand data.
For example, Wei and Chen [8] used empirical mode decomposition (EMD) to obtain the intrinsic mode function (IMF) components of passenger flow data and identified the useful IMFs as inputs for the back-propagation neural network (BPN); Sun et al. [9] extracted the sub-sequences of demand series using wavelet transformation and then applied the support vector machine (SVM) to forecast the extracted sub-sequences. The final demand predictions were reconstructed using the wavelet inverse transformation. Ma et al. [2] constructed weekly, daily, and hourly pattern time series, and developed an interactive multiple model-based pattern hybrid (IMMPH) approach to dynamically select the best model (matching the time series pattern) to predict passenger demand. Recently, the deep learning-based prediction framework gains widely applications. In this area, the stacked autoencoder (SAE) technique is frequently used in extracting spatiotemporal features [3], [10]. For example, Ly et al. [3] devised an SAE model to extract spatiotemporal features of traffic flow. A logistic regression function is utilized to combine the extracted features to predict the traffic demand. Meanwhile, the convolution neural network (CNN) has been widely applied to capture spatiotemporal dependencies of multiple time series [11], [12]. For instance, Ma et al. [13] proposed a parallel architecture that includes the CNN and bi-directional long short-term memory networks (BLSTM) to extract spatiotemporal features of passenger demand in large-scale metro networks. The model performance is better than statistical models (e.g., ARIMA) and deep learning models (e.g., CNN-BLSTM).

The prediction functions include: a) statistic models, such as ARIMA [14], state-space model [15], and Bayesian networks [16], b) machine learning models, such as SVM [17], Kalman filtering [18], and shallow neural networks [19] and c) deep learning models. A review of prediction methods can refer to [20], [21]. Deep learning models have gained increasing interest for its extraordinary mapping ability for complex data. For example, Hao et al. [22] constructed a sequence to sequence model embedded with the attention mechanism to predict alighting passengers in a large-scale metro system; Polson and Sokolov [23] developed a deep neural network (DNN) to predict traffic flow under abnormal conditions; Tsai et al. [24] combined simulated annealing (SA) algorithm and DNN to predict bus passenger demand; Zhang et al. [25] utilized a spatial-temporal graph inception residual network to predict the network-based traffic flow. Deep belief network (DBN) [26], LSTM NN [27], radial basis function networks (RBFNN) [28] were also reported in literature. Although deep learning is popular in predictions, many studies have provided evidence on supporting simple statistical or machine learning models [29], [30]. Developing appropriate functions to match the time series patterns is the key to effective predictions.

From the perspective of prediction outputs, many methods were proposed to combine the sub-model outputs by taking advantage of real-time observations. The hybrid model yielded better prediction than the individual model by selecting the best ‘fit’ model predictions with real-time observations [20]. Diao et al. [30] combined the outputs of an efficient tracking model and a novel Gaussian process model to predict the real-time passenger demand, and the ensemble prediction can improve the accuracy by 20%-50% on average. Ding et al. [31] integrated the outputs of ARIMA and generalized autoregressive conditional heteroskedasticity (GARCH) to estimate the volatility of the passenger demand. Also, Gu et al. [32] used the Bayesian method to combine the outputs of the three sub-predictors, including the gated recurrent unit neural network (GRU NN), ARIMA, and radial basis function neural network (RBFNN). Noursalehi et al. [33] combined the prediction of the state-space model and dynamic factor model, which shows promising performance for abnormal passenger flow prediction.

Time series data exhibit significant characteristics. Therefore, identifying the characteristics (through transformation or decomposition) have been proven to improve the forecasting accuracy in areas of the economy [34], [35], atmospheric science [35], [36], and transportation [8], [9], [35]. For example, considering the periodicity of time series data, Chen et al. [35] proposed a Periodicity-based Parallel Time Series Prediction (PPTSP) algorithm for large-scale time series data prediction. Compressing the data and extracting the feature accordingly, a multi-layer time series periodic pattern recognition (MTSPPR) algorithm using the Fourier Spectrum Analysis (FSA) is developed to identify the potential multi-layer periodicity. A Periodicity-based Time Series Prediction (PTSP) algorithm is proposed to make predictions. Given various time series data (e.g., meteorological datasets, traffic flow datasets, and stock price datasets), results showed that the PPTSP algorithm significantly outperformed other algorithms regarding prediction accuracy and performance. Similarly, Yang et al. [37] analyzed the influence of periodic components on short-term speed prediction. Using the periodic component of speed data as the model input, results validated that combining the periodic component with statistical models (e.g., ARIMA) or machine learning models (e.g., SVM) could improve the multi-step ahead prediction accuracy. However, to the best of our knowledge, the similar idea was only carried out by the EMD [8] and wavelet [9] technique in short-term metro passenger flow prediction. Unfortunately, the EMD or wavelet decompositions could hardly represent passenger demand patterns in urban railway systems and makes it difficult for prediction models matching these low interpretability patterns. Besides, the predictability of passenger demand patterns should be the prerequisite for choosing suitable models, which is neglected in the previous studies [8], [9]. Therefore, the paper bridges the knowledge gap by exploring the potential of the decomposition-based prediction idea for urban railway demand and deriving insights on aspects of decomposition methods, one- and multi-step predictions, and station/passenger types selections.

In the paper, the Seasonal and Trend decomposition using Loess (STL) [38], Holt-Winters (HW) method [40], [41],
and LSTM NN [44] are proposed (i.e., STL-HW-LSTM) to predict station-based short-term metro passenger flow under one- and multi-step ahead prediction scenarios. Preprocessing passenger flow sequences by STL, we obtain the seasonal, trend, and residuals accordingly [38], which represents the constant, long-term fluctuant, and stochastic passenger demand patterns in the context of short-term metro passenger flow prediction. Then, approximate entropy (ApEn) is employed to quantify the predictability over each subsequence. After that, the HW method, which is effective in predicting time series with high regularity, is employed to predict the component with high predictability (i.e., constant and long-term fluctuant passenger demand). LSTM NN, which shows a strong mapping ability in time series predictions, is utilized to predict the component with low predictability (i.e., stochastic passenger demand). Combining the predictions from the HW method and LSTM NN via STL and ApEn, the proposed model can provide more reliable and robust predictions under both one- and multi-step ahead prediction scenarios.

The main contributions of this work are three-fold:

- The STL decomposition-based hybrid model outperforms other representative models for both one-step and multi-step predictions in terms of accuracy and robustness.
- ApEn is adopted to evaluate the predictability of passenger flow time series and lays the basis for choosing suitable prediction models.
- To assess the impact of target stations and passenger flow types on the model prediction performance, we employ the AHC algorithm to select target stations, and consider the entry flow, exit flow, and transfer flow.

The rest of the paper is organized as follows. Section 2 describes STL, ApEn, the HW method, LSTM NN, and the proposed model. Section 3 demonstrates the data and settings for the case study. Section 4 presents the experiment results and the findings. Section 5 concludes the paper.

II. METHODOLOGY

Fig. 1 shows the proposed hybrid STL-HW-LSTM method for the station-based short-term urban rail passenger flow prediction. The inputs are passenger demand time series of stations (e.g., extracted from smart card data), and the outputs are passenger demand predictions in the future time intervals (e.g., 15 minutes ahead). The proposed model consists of four stages:

- **Stage 1**: time series decomposition. STL decomposes the passenger flow time series into seasonal, trend, and residual time series. The seasonal, trend, and residual time series represent stable, long-term fluctuant, and stochastic passenger demand, respectively.
- **Stage 2**: predictability quantification. ApEn is utilized to evaluate the predictability of stable, long-term fluctuant, and stochastic passenger demand to lay the basis for choosing suitable for each subsequence.
- **Stage 3**: time series predictions. It selects and builds corresponding functions to predict the seasonal, trend and residual time series. Particularly, HW models are developed to predict the seasonal and trend time series given their high regularity. LSTM NN model is used to predict the residuals that show highly irregularity and nonlinear correlation structures over time.
- **Stage 4**: prediction hybrid. It combines the predictions from seasonal, trend and residual models to predict the passenger demand in the future steps.

A. SEASONAL AND TREND DECOMPOSITION USING LOESS (STL)

STL was initially introduced by Robert *et al.* [38]. It decomposes time series by applying a series of smoothing approaches using Locally Weighted Regression-Loess. Given a time series $X_t$, STL decomposes $X_t$ into three additive components containing seasonality $S_t$, trend $T_t$ and remainder $R_t$:

$$X_t = S_t + T_t + R_t$$  \hspace{1cm} (1)

where $t$ represents the time interval (e.g., 7:00-7:15). STL is robust to outliers. This can be beneficial to real-world applications where missing and noisy observations are inevitable. In urban rail system, the travel characteristics (e.g., departure and arrival time, origins, and destinations) among passengers are varied, which results in a stable and stochastic passenger demand pattern, respectively. Therefore, separating the seasonal and random component effectively, STL instead of other decomposition methods is adopted. The prediction task...
is then transformed into the prediction of different passenger demand patterns.

STL consists of inner and outer recursive procedures. The inner loop extracts the seasonal and trend patterns, and the outer loop calculates the residual. Take the \( (k + 1) \) th inner loop as an example, the updates of time series \( X_{t}^{k+1} \) and \( T_{t}^{k+1} \) are:

Step 1: **Detrending.** Compute the detrended series, \( \hat{X}_{t}^{k+1} = X_{t} - T_{t}^{k} \).

Step 2: **Cycle-subseries Smoothing.** Calculate the temporary seasonal series \( C_{t}^{k+1} \) by smoothing each cycle-subseries of \( \hat{X}_{t}^{k+1} \) using loess smoother. For a monthly series over a year, January series is the first cycle-subseries, February second, and so on so forth.

Step 3: **Low-pass Filtering of Smoothed Cycle-Subseries.** Estimate temporary trend components \( \hat{T}_{t}^{k+1} \) by applying moving average filters and loess smoother to \( C_{t}^{k+1} \).

Step 4: **De-trending of Smoothed Cycle-Subseries.** Extract the seasonal component \( S_{t}^{k+1} \) by subtracting \( \hat{T}_{t}^{k+1} \) from \( X_{t} \).

Step 5: **De-seasonalizing.** Construct the de-seasonalized series \( \hat{X}_{t}^{season} = X_{t} - S_{t}^{k+1} \).

Step 6: **Trend Smoothing.** Compute the trend component \( T_{t}^{k+1} \) by smoothing \( \hat{X}_{t}^{season} \) using the loess smoother.

In the outer loop, calculate the residual component by subtracting the seasonal and trend components (i.e., \( S_{t}^{k+1} \) and \( T_{t}^{k+1} \)) from the original passenger flow time series \( X_{t} \).

The robustness weight is defined for each data to measure the reliability of the corresponding \( R_{t} \). The weight will be used in the next inner iteration (i.e., the loess function of steps 2 and 6) to weaken the effects of outliers. Readers can refer to [38] for the form of loess function and the selection of other parameters.

### B. APPROXIMATE ENTROPY (APEN)

Given passenger demand time series that obtained from STL, choosing the suitable prediction model is a key issue. Therefore, we attempt to evaluate the regularity and predictability of passenger demand time series at first and then select the appropriate models accordingly. Approximate entropy (ApEn) technique was developed by Pincus [39] used to quantify the amount of regularity and predictability over time series. Accurate entropy calculation requires vast amounts of data and the results will be greatly influenced by system noise [39], thus it is not practical to apply these methods to experimental data. ApEn can handle these limitations by modifying an exact regularity statistic. The steps are:

Step 1: Fix an integer \( m \) and a positive real number \( r \) for a time series \( X_{t} = (x_{1}, x_{2}, \ldots, x_{N}) \), and \( m \) represents the length of compared run of \( X_{t} \) and a filtering level, respectively. \( N \) is the length of \( X_{t} \).

Step 2: Form a sequence of vectors \( U_{1}, U_{2}, \ldots, U_{N-m+1} \) in the \( m \)-dimensional space: \( U_{i} = [x_{i}, x_{i+1}, \ldots, x_{i+m-1}] \).

Step 3: For each \( i \) (\( 1 \leq i \leq N - m + 1 \)), use the sequence \( U_{1}, U_{2}, \ldots, U_{N-m+1} \) to construct:

\[
C_{i}^{m} (r) = \frac{n_{i}}{N - m + 1}
\]

\( n_{i} \) is the number of \( x_{j} \) that satisfies \( d \left[ x_{i}, x_{j} \right] \leq r \):

\[
d \left[ x_{i}, x_{j} \right] = \max_{a} \left| x_{a} - x_{j} \right|
\]

The \( x_{i} \) are the \( m \) scalar components of \( U \).

Step 4: Define ApEn as:

\[
\text{ApEn} = \Phi^{m} (r) - \Phi^{m+1} (r)
\]

\[
\Phi^{m} (r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \log \left( C_{i}^{m} (r) \right)
\]

ApEn reflects the likelihood that similar patterns of observations will not be followed by additional similar observations. A time series containing many repetitive patterns has a relatively small ApEn, which owes a higher predictability.

The ApEn provides a way to quantify the predictability of the decomposed time series for selecting the appropriate prediction models, especially for cases when the predictability of the decomposed time series is ambiguous (e.g., MED and wavelet).

### C. HOLT-WINTERS (HW)

Passenger flow time series exhibits strong seasonal patterns. The HW method, proposed by Holt [40] and Winters [41], is adopted to forecast the seasonal and trend components. The approach belongs to a class of exponential smoothing techniques that aim to predict the time series \( X_{t} \) by dividing \( x_{t} \in X_{t} \), into level \( l_{t} \), seasonal \( s_{t} \), and trend terms \( b_{t} \). HW method has a solid theoretical merit on prediction accuracy [43].

The additive HW is defined as:

\[
l_{t} = \alpha (x_{t} - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1})
\]

\[
b_{t} = \beta (l_{t} - l_{t-1}) + (1-\beta)b_{t-1}
\]

\[
s_{t} = \gamma (y_{t} - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}
\]

\[
\hat{y}_{t+h} = l_{t} + h b_{t} + s_{t-m+h}
\]

The multiplicative HW is given by:

\[
l_{t} = \alpha \frac{x_{t}}{s_{t-m}} + (1-\alpha)(l_{t-1} + b_{t-1})
\]

\[
b_{t} = \beta \left( \frac{l_{t}}{l_{t-1}} - 1 \right) + (1-\beta)b_{t-1}
\]

\[
s_{t} = \gamma \frac{x_{t}}{l_{t-1} - b_{t-1}} + (1-\gamma)s_{t-m}
\]

\[
\hat{y}_{t+h} = \left( l_{t} + h b_{t} \right) s_{t-m+h}
\]

where \( h \) represents the forecasting horizon (i.e., forecasting step), \( \alpha \), \( \beta \), and \( \gamma \) are the smoothing parameters.

### D. LONG SHORT-TERM MEMORY NETWORKS (LSTM NN)

LSTM NN solve the vanishing gradient problem that occurs in the traditional recurrent neural networks (RNN) [44]. Fig. 2 shows the basic structure of LSTM NN. A standard LSTM NN structure is composed of one input layer, one recurrent hidden layer, and one output layer. The core of the hidden layer is a memory block, consisting of a memory cell, and input, output gate, and forget gates. The cell transports values at arbitrary time intervals and memorizes the temporal state. These gates serve for similar purposes with...
the “conventional” artificial neurons in the feedforward NN. Through the multiplicative gates, the memory cells can store and access information over a long time period.

Given the historical time series \( x = (x_1, x_2, \cdots, x_T) \), the target time series \( y = (y_1, y_2, \cdots, y_T) \), and the hidden output \( h = (h_1, h_2, \cdots, h_T) \), the prediction process is iteratively carried out by equations (14)-(20):

\[
i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \\
f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \\
c_t = f_t * c_{t-1} + i_t * \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\
o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{18} \\
h_t = o_t * \tanh(c_t) \tag{19} \\
y_t = \sigma(W_{yh}h_t + b_y) \tag{20}
\]

where \( T \) is the prediction period, \( t \) is the time interval, \( W \) is the weight matrix (e.g., \( W_{xi} \) is the weight matrix between input \( x \) and input gate \( i \)), and \( b \) is the bias vector (e.g., \( b_i \) is the bias vector of input gate). \( i, f, o, \) and \( c \) represent the input gate, forget gate, output gate, and cell activation vectors, respectively. \( \Delta(\cdot) \) is the standard logistic sigmoid function:

\[
\sigma(x) = \frac{1}{1 + e^{-x}} \tag{20}
\]

III. EXPERIMENT SETUP

A. DATA

We use the AFC data from the Shanghai Metro. It covers all stations over three months from July 1st to September 30th, 2016. Fig. 3 depicts the Shanghai Metro network. By 2016, there are 14 metro lines and 313 stations. Metro line 1 (red color in Fig. 3) is one of the busiest lines in Shanghai and is selected in the case study. Table 1 lists the AFC data fields used in this paper. As suggested in [45], 15 minutes is a threshold when making short-term metro passenger flow prediction. Therefore, we aggregate the station entry/exit flow into 15 minutes for the prediction analysis. The prediction time period is the common operation time 6:00-23:00, which results in 68 data points per day. The data from the last week is used for testing, and the remaining is for training.

The mean absolute percentage error (MAPE) and root mean squared error (RMSE) are used as performance metrics. The former measures the mean prediction accuracy while the latter reflects the forecasting stability.

\[
\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \tag{21}
\]

\[
\text{RMSE} = \left[ \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \right]^{0.5} \tag{22}
\]

B. TARGET STATION SELECTION

The distribution of metro station entry/exit flows over time may vary among stations. To evaluate the model performance, we employ the AHC algorithm to cluster stations based on entry/exit demand patterns. Fig. 4 shows an example of clustering results for metro line 1. The detailed information on AHC algorithm can refer to [33].

Fig. 5 shows the demand pattern for three groups of stations, including (a) peak demand in the evening; (b) peak demand in the morning; (c) peak demand in both
The residuals, and in the STL-HW-LSTM, the LSTM NN is fully tailored to each application [38]. After exploring variables are straight, and the last parameter, \( n_t \), should be carefully tailored to each application [38]. After exploring various combinations, the following STL parameters are used: \( n_p = 476, n_l = 2, n_o = 10, n_t = 476, n_s = 477, \) and \( n_i = 717. \)

The length of compared run of time series \( X_t \) and a filtering level (i.e., \( m \) and \( r \)) are prerequisites in the application of ApEn. Yentes et al. [47] conducted a comprehensive study on the appropriate use of ApEn with short data. When the length of a dataset is greater than 200 (i.e., the case in the paper), 2 is suitable for \( m \) and several \( r \) values should be examined. Attempting the empirical values of \( r \) (i.e., 0.1, 0.2, 0.3), we verified that 0.2 was the optimal value in the experiment.

The additive HW is used to model the seasonal time series. We use the maximum likelihood estimation [47] to calculate the smoothing parameters of level \( \alpha \), trend \( \beta \), and seasonality \( \gamma \). Table 2 shows the resulting parameters.

In the STL-HW model, we build the HW method to predict the residuals, and in the STL-HW-LSTM, the LSTM NN is developed to predict residuals. Time series aggregated in four adjacent time intervals [46], [49], [50] is used to predict the passenger demand in the next time slot. Hence, the model input is \( x_{t-4}, x_{t-3}, x_{t-2}, x_{t-1} \), and the output is \( x_t \).

We employ the grid search and cross-validation technique employed to select the optimal parameters in LSTM NN. Table 3 shows the optimal parameter values for LSTM NN.

The ratio of the validation dataset is set to be 0.2. The value of cross-validation is empirically set as 3~10 due to the massive calculation, and we set \( k \) as 5. The hidden layer size ranges from 1 to 6 [3]–[10], the number of hidden units \( l_N \in \{8, 12, 16, 20, 24\} \), the activation function of each layer is “tanh”, the batch-size \( l_S \in \{1, 2, 3, 4\} \), the learning rate \( l_L \in \{0.01, 0.05, 0.001, 0.005, 0.0001\} \), and a dense layer is added to the output layer at last. Table 3 shows the optimal values of LSTM NN:

### Table 2. The parameter values of HW.

| Station | Passenger type | Seasonal | Trend s | Residual |
|---------|----------------|----------|---------|----------|
| XZ      | Entry flow     | \( \alpha \) = 0.197, \( \beta \) = 0.726, \( \gamma \) = 0.371 | 0.000, 0.055 |
| CBL     | Exit flow      | 0.053, 0.053, 0.0007, 0.035, 0.021, 0.032 | 0.000, 0.023 |
| RMGC    | Transfer       | 0.466, 0.060, 0.0406, 0.0003, 0.0007 | 0.079, 0.000, 0.013 |

The parameter values of LSTM NN.

| Hidden layers | Hidden neurons | Learning rate | Batch size |
|---------------|----------------|---------------|------------|
| 1             | 20             | 0.0001        | 1          |

The parameter values of LSTM NN.

The parameter values of ARIMA.

For BPNN, the neurons in the input layer, hidden layer, and output layer are 4, 4, and 1, such 4*4*1 structure is also used by [46], [49], [50].

3) ND-LSTM

The naive decomposition uses derivation by subtracting the average passenger flow from the observed one for the same
time period of a day. The residual passenger flow is:

\[ \bar{r}_t = X_t - \bar{x}_t \]

where \( \bar{x} \) is the average passengers flow at time period \( t \) for the same weekdays or weekends. \( \bar{x} \) is constant in the studied dataset. LSTM NN is used to predict the residual time series. The determination of the optimal parameters of LSTM NN is similar to that in STL-HW-LSTM, and the results are the same as shown in Table 3. The optimal parameters of LSTM NN are the same as STL-HW-LSTM (Table 3).

IV. RESULTS AND DISCUSSION

A. ENTRY FLOW PREDICTION AT XZ STATION

Fig. 6 shows the STL decomposition of entry flow at XZ station from Sept. 23 to 30, 2016. The seasonal component is a smoothing result of the passenger flow time series, thus the outliers are removed from the raw data. The trend passenger demand is almost stable, and the stochastic passenger demand exhibits weak regularity. Table 5 shows the ApEn value of each passenger flow time series.

Fig. 7 shows the multi-step ahead (i.e., \( h \)) prediction deviations of entry flows. Table 6 shows that STL-HW-LSTM, STL-HW, and ND-LSTM models significantly outperform the ARIMA-BPNN model, especially for the multi-step predictions. The STL-HW-LSTM model performs the best for all scenarios in both MAPE and RMSE. The MAPE for the STL-HW-LSTM model is less than 8% in the up to four-step ahead predictions.

B. EXIT PASSENGER FLOW PREDICTION AT CBL STATION

Fig. 8 shows the prediction results of exit passenger flow. Table 7 presents the ApEn of each exit passenger demand pattern at CBL station.

Table 8 shows that the STL-HW-LSTM model outperforms other models in multi-step predictions. The MAPEs of all the models in Table 8 are higher than those in Table 6. It attributes...
TABLE 7. ApEn of entry passenger flow at CBL station.

|          | Seasonal component | Trend component | Residual component |
|----------|-------------------|----------------|-------------------|
| ApEn     | 0.429             | 0.036          | 1.256             |

TABLE 8. Forecasting errors of exit flow passenger flow predictions at CBL station.

| Forecasting horizon | STL-HW-LSTM | STL-HW | ARIMA-BPNN | ND-LSTM |
|---------------------|-------------|--------|------------|---------|
| MAPE                | RMSE        | MAPE   | RMSE       | MAPE    | RMSE |
| 15-min              | 11.75       | 28.51  | 11.77      | 36.60   | 14.96 | 38.46 | 11.49 | 30.42 |
| 30-min              | 12.35       | 29.19  | 13.77      | 32.07   | 23.04 | 59.17 | 15.24 | 32.02 |
| 45-min              | 12.22       | 29.35  | 13.26      | 32.58   | 30.23 | 85.88 | 13.40 | 33.14 |
| 60-min              | 12.58       | 31.60  | 15.85      | 35.91   | 36.10 | 102.39| 13.09 | 36.12 |

The reason is that the predictability of transfer flow is reduced due to the operation.

D. DISCUSSION

The case of transfer passenger flow prediction is analyzed because it best supports the predictive ability of the proposed model. Fig. 10 shows the prediction of each component. Table 11 summarizes the corresponding forecasting errors.

TABLE 10. Forecasting errors of transfer passenger flow at RMGC station.

| Forecasting horizon | STL-HW-LSTM | STL-HW | ARIMA-BPNN | ND-LSTM |
|---------------------|-------------|--------|------------|---------|
| MAPE                | RMSE        | MAPE   | RMSE       | MAPE    | RMSE |
| 15-min              | 15.37       | 27.22  | 18.00      | 32.03   | 35.02 | 42.26 | 17.53 | 29.32 |
| 30-min              | 15.14       | 26.34  | 18.79      | 31.52   | 48.12 | 68.27 | 17.62 | 29.98 |
| 45-min              | 14.87       | 26.03  | 18.52      | 30.55   | 118.74| 138.48| 17.74 | 30.58 |
| 60-min              | 15.19       | 26.93  | 20.46      | 33.54   | 126.14| 200.65| 18.12 | 32.46 |

With the HW method, the value of MAPE and RMSE of seasonal and trend components are small in Table 11. However, the method cannot adequately predict the residuals with high noise, and thus the corresponding MAPE is at least 47.57%. In contrast, LSTM NN reduces the MAPE of residuals by at least 2.47% compared with the HW method. Therefore, it is safe to conclude that the proposed model outperforms the STL-HW.

For the ND-LSTM, simply averaging passenger flow time series can not separate the trend and residual components and leads to the fact that the LSTM NN cannot accurately model the trend component. Hence, the proposed model also performs better than the ND-LSTM. It needs to note that the most ND-based residuals in Fig. 10 are between 0 to 10, and a minor prediction error can generate a tremendous value for MAPE (Table 11).

Additionally, the forecasting errors of the seasonal and trend components barely change, although the forecasting step is increasing. It is because the seasonal and trend passenger demand remains constant in a short-time span (e.g., three months). Therefore, for the STL-HW-based models (i.e., STL-HW-LSTM and STL-HW), we only focus on the stochastic passenger demand prediction, and thus the...
prediction accuracy can be significantly improved. (Tables 6, 8, and 10)

The experiment results strongly support that STL is promising for the multi-step predictions by providing regular and random passenger demands separately for the forecasting model. By calculating the predictability for each decomposed passenger flow time series, we prove that ApEn is an efficient tool to help us select suitable prediction models. Combining the prediction results of HW and LSTM NN, the predictive advantages of these two models are combined. Therefore, the proposed model has great potential to provide more robust and reliable forecasting abilities.

V. CONCLUSION

The paper proposes a STL-based short-term metro passenger flow prediction model. It decomposes the original passenger flow into constant, long-term fluctuant and stochastic time series. We use the ApEn to evaluate the predictability of each time series, which lays the basis for choosing the suitable prediction models. We build the HW models to predict the constant and long-term fluctuant time series (high regularity), and develop the LSTM NN model to model the stochastic time series (irregularity).

Case studies using a heavily used urban railway system validated the better accuracy performance of the proposed model STL-HW-LSTM in predicting entry/exit/transfer flows, compared with STL-HW[42], ARIMA-BPNN[46], and a (derivation) decomposition-based LSTM NN. Also, the proposed STL-HW-LSTM is significantly more robust (2% in MAPE) in multi-step prediction than the peer models.

REFERENCES

[1] H. N. Koutsopoulos, Z. Ma, P. Noursalehi, and Y. Zhu, “Transit data analytics for planning, monitoring, control, and information,” in Mobility Patterns, Big Data and Transport Analytics. Amsterdam, The Netherlands: Elsevier, Jan. 2019, pp. 259–261.

[2] Z. Ma, J. Xing, M. Mesbah, and L. Ferreira, “Predicting short-term bus passenger demand using a pattern hybrid approach,” Transp. Res. C, Emerg. Technol., vol. 39, pp. 148–163, Feb. 2014.

[3] 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.

[4] J. W. C. van Lint, S. P. Hoogendoorn, and H. J. van Zuylen, “Accurate freeway travel time prediction with state-space neural networks under missing data,” Transp. Res. C, Emerg. Technol., vol. 13, nos. 5–6, pp. 347–369, Oct. 2005.

[5] E. Jenelius, “Data-driven metro train crowding prediction based on real-time load data,” IEEE Trans. Intell. Transp. Syst., vol. 21, no. 6, pp. 2254–2265, Jun. 2020.

[6] F. C. Pereira, F. Rodrigues, and M. Ben-Akiva, “Using data from the Web to predict public transport arrivals under special events scenarios,” J. Intell. Transp. Syst., vol. 19, no. 3, pp. 273–288, Jul. 2015.

[7] M. Ni, Q. He, and J. Gao, “Forecasting the subway passenger flow under event occurrences with social media,” IEEE Trans. Intell. Transp. Syst., vol. 18, no. 6, pp. 1623–1632, Jun. 2017.

[8] Y. Wei and M.-C. Chen, “Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks,” Transp. Res. C, Emerg. Technol., vol. 21, no. 1, pp. 148–162, Apr. 2012.

[9] 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.

[10] 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.

[11] J. Ke, H. Zheng, H. Yang, and X. M. Chen, “Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach,” Transp. Res. C, Emerg. Technol., vol. 85, pp. 591–608, Dec. 2017.

[12] B. Du, H. Peng, S. Wang, M. Z. A. Bhuiyan, L. Wang, Q. Gong, L. Liu, and J. Li, “Deep irregular convolutional residual LSTM for urban traffic passenger flows prediction,” IEEE Trans. Intell. Transp. Syst., vol. 21, no. 3, pp. 972–985, Mar. 2020.

[13] 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.

[14] B. M. Williams and L. A. Hoel, “Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results,” J. Transp. Eng., vol. 129, no. 6, pp. 664–672, Nov. 2003.

[15] A. Stathopoulos and M. G. Karlaftis, “A multivariate state space approach for urban traffic flow modeling and prediction,” Transp. Res. C, Emerg. Technol., vol. 11, no. 2, pp. 121–135, Apr. 2003.
[16] E. Castillo, J. M. Menéndez, and S. Sánchez-Cambroneró, “Predicting traffic flow using Bayesian networks,” Transp. Res. B, Methodol., vol. 42, no. 5, pp. 482–509, Jun. 2008.

[17] N. Sapankevych and R. Sankar, “Time series prediction using support vector machines: A survey,” IEEE Comput. Intell. Mag., vol. 4, no. 2, pp. 24–38, May 2009.

[18] I. Okutani and Y. J. Stephanead, “Dynamic prediction of traffic volume through Kalman filtering theory,” Transp. Res. B, Methodol., vol. 18, no. 1, pp. 1–11, Feb. 1984.

[19] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golia, “Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach,” Transp. Res. C, Emerg. Technol., vol. 13, no. 3, pp. 211–234, Jun. 2005.

[20] E. I. Vlahogianni, J. C. Golia, and M. G. Karlaftis, “Short-term traffic forecasting: Overview of objectives and methods,” Transp. Rev., vol. 24, no. 5, pp. 533–557, Sep. 2004.

[21] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golia, “Short-term traffic forecasting: Where we are and where we’re going,” Transp. Res. C, Emerg. Technol., vol. 43, pp. 3–19, Jun. 2014.

[22] S. Hao, D.-H. Lee, and D. Zhao, “Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system,” Transp. Res. C, Emerg. Technol., vol. 107, pp. 287–300, Oct. 2019.

[23] N. G. Polson and V. O. Sokolov, “Deep learning for short-term traffic flow prediction,” Transp. Res. C, Emerg. Technol., vol. 79, pp. 1–17, Jun. 2017.

[24] C.-W. Tsai, C.-H. Hsia, S.-J. Yang, S.-J. Liu, and Z.-Y. Fang, “Optimizing hyperparameters of deep learning in predicting bus passengers based on simulated annealing,” Appl. Soft Comput., vol. 88, Mar. 2020, Art. no. 106068.

[25] Y. Zhang, T. Cheng, and Y. Ren, “A graph deep learning method for short-term traffic forecasting on large road networks,” Comput.-Aided Civ. Infrastruct. Eng., vol. 34, no. 10, pp. 877–896, Oct. 2019.

[26] W. Huang, G. Song, H. Hong, and K. Xie, “Deep architecture for traffic flow prediction: Deep belief networks with multitask learning,” IEEE Trans. Intell. Transp. Syst., vol. 15, no. 5, pp. 2191–2201, Oct. 2014.

[27] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, “Long-term memory neural network for traffic speed prediction using remote microwave sensor data,” Transp. Res. C, Emerg. Technol., vol. 54, pp. 187–197, May 2015.

[28] Y. Li, X. Wang, 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.

[29] L. Tang, Y. Zhao, J. Cabrera, J. Ma, and K. L. Tsui, “Forecasting short-term passenger flow: An empirical study on shenzhen metro,” IEEE Trans. Intell. Transp. Syst., vol. 20, no. 10, pp. 3613–3622, Oct. 2019.

[30] Z. Diao, D. Zhang, X. Wang, K. Xie, S. He, X. Lu, and Y. Li, “A hybrid model for short-term traffic volume prediction in massive transportation systems,” IEEE Trans. Intell. Transp. Syst., vol. 20, no. 3, pp. 935–946, Mar. 2019.

[31] 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.

[32] Y. Gu, W. Lu, X. Xu, L. Qin, Z. Shao, and H. Zhang, “An improved Bayesian combination model for short-term traffic prediction with deep learning,” IEEE Trans. Intell. Transp. Syst., vol. 21, no. 3, pp. 1332–1342, Mar. 2020.

[33] P. Noursalehi, H. N. Koutsopoulos, and J. Zhao, “Real-time transit demand prediction capturing station interactions and impact of special events,” Transp. Res. C, Emerg. Technol., vol. 97, pp. 277–300, Dec. 2018.

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

[35] J. Chen, K. Li, H. Rong, K. Bilal, K. Li, and P. S. Yu, “A periodicity-based parallel time series prediction algorithm in cloud computing environments,” Inf. Sci., vol. 496, pp. 506–537, Sep. 2019, doi: 10.1016/j.ins.2018.06.045.

[36] V. Nourani, M. T. Alami, and M. H. Aminfar, “A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation,” Eng. Appl. Artif. Intell., vol. 22, no. 3, pp. 466–472, Apr. 2009.

[37] X. Yang, Y. Zou, J. Tang, J. Liang, and M. Ijaz, “Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models,” J. Adv. Transp., vol. 2020, pp. 1–16, Jun. 2020, doi: 10.1155/2020/9628957.

[38] C. Robert, C. William, and T. Irna, “STL: A seasonal-trend decomposition procedure based on loess,” J. Off. Statist., vol. 6, no. 1, pp. 3–73, 1990.

[39] S. M. Pincus, “Approximate entropy as a measure of system complexity,” Proc. Nat. Acad. Sci. USA, vol. 88, no. 6, pp. 2297–2301, Mar. 1991.

[40] C. C. Holt, “Forecasting trends and seasonal by exponentially weighted averages,” Carnegie Inst. Technol., Pittsburgh, PA, USA, ONR Memorandum no. 52, 1957.

[41] P. R. Winters, “Forecasting sales by exponentially weighted moving average,” Manage. Sci., vol. 6, no. 3, pp. 324–342, Apr. 1960.

[42] T. M. Danas, F. L. Cyriano Oliveira, and H. M. Varela Repolho, “Air transportation demand forecast through bagging Holt winters method,” J. Air Transp. Manage., vol. 59, pp. 116–123, Mar. 2017.

[43] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose, “A state space framework for automatic forecasting using exponential smoothing methods,” Int. J. Forecasting, vol. 18, no. 3, pp. 439–454, Jul. 2002.

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

[45] C. Zhong, M. Batty, E. Manley, J. Wang, Z. Wang, F. Chen, and G. Schmitt, “Variability in regularity: Mining temporal mobility patterns in London, Singapore and Beijing using smart-card data,” PLoS ONE, vol. 11, no. 2, Feb. 2016, Art. no. e0149222, doi: 10.1371/journal.pone.0149222.

[46] G. P. Zhang, “Time series forecasting using a hybrid ARIMA and neural network model,” Neurocomputing, vol. 50, pp. 159–175, Jan. 2003.

[47] J. M. Yentes, N. Hunt, K. K. Schmid, J. P. Kaipust, D. McGrath, and N. Stergiou, “The appropriate use of approximate entropy and sample entropy with short data sets,” Ann. Biomed. Eng., vol. 41, no. 2, pp. 349–365, 2013.

[48] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 2nd ed. Melbourne, VIC, Australia: OTexts, 2018, pp. 207–213.

[49] M. Cottrell, B. Girard, Y. Giraud, M. Mangaes, and C. Muller, “Neural modeling for time series: A statistical stepwise method for weight elimination,” IEEE Trans. Neural Netw., vol. 6, no. 6, pp. 1355–1364, Nov. 1995.

[50] C. de Groot and D. Würtz, “Analysis of univariate time series with connectionist nets: A case study of two classical examples,” Neurocomputing, vol. 3, no. 4, pp. 177–192, Nov. 1991.

[51] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis, Forecasting and Control, 5th ed. Hoboken, NJ, USA: Wiley, 2015, pp. 179–194.
YI YANG received the B.E. degree in transportation and the M.E. degree in traffic and transportation engineering from Central South University, Changsha, China, in 2014 and 2017, respectively. He is currently pursuing the Ph.D. degree in transportation planning and management with the School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, Sichuan, China. His research interests include public transportation and urban transit system operation and management.

WENHUA JIANG received the B.E. degree in traffic engineering from the Suzhou University of Science and Technology, Suzhou, China, the M.E. degree in transportation engineering from Tongji University, Shanghai, China. She is currently pursuing the Ph.D. degree with Monash University, Melbourne, VIC, Australia. Her research interests include spatio-temporal data processing, machine learning, and urban transit system operation.

XINGUO JIANG received the B.E. and M.E. degrees in transportation engineering from Tongji University, Shanghai, China, and the Ph.D. degree in transportation engineering from Michigan State University. He was with International Traffic Consulting Firm and City of Fort Myers, Florida, as a Chief Traffic Engineer. He is currently a Professor with the School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, Sichuan, China. His research interests include traffic planning and management, public transportation, and traffic safety. He received the Honorable Title of Sichuan 100 Plan, in 2012.