IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 19, 2019, accepted August 18, 2019, date of publication August 22, 2019, date of current version September 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2936938

Wavelet-HST: A Wavelet-Based Higher-Order Spatio-Temporal Framework for Urban Traffic Speed Prediction

NA ZHANG, XUEFENG GUAN, JUN CAO, XINGLEI WANG, AND HUAYI WU
State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan 430079, China

Corresponding author: Xuefeng Guan (guanxuefeng@whu.edu.cn)

This work was supported by the National Key Research and Development Program of China under Grant 2017YFB0503802 and the National Natural Science Foundation of China program 41971348.

ABSTRACT As a crucial part of the Intelligent Transportation System, traffic forecasting is of great help for traffic management and guidance. However, predicting short-term traffic conditions on a large-scale road network is challenging due to the complex spatio-temporal dependencies found in traffic data. Previous studies used Euclidean proximity or topological adjacency to explore the spatial correlation of traffic flows, but did not consider the higher-order connectivity patterns exhibited in a road network, which have a significant influence on traffic propagation. Meanwhile, traffic sequences display distinct multiple time-frequency properties, yet few researchers have made full use of this resource. To fill this gap, we propose a novel hybrid framework – Wavelet-based Higher-order Spatial-Temporal method (Wavelet-HST) to accurately predict network-scale traffic speeds. Wavelet-HST first uses discrete wavelet transform (DWT) to decompose raw traffic data into several components with different frequency sub-bands. Then a motif-based graph convolutional recurrent neural network (Motif-GCRNN) is proposed to learn the higher-order spatio-temporal dependencies of traffic speeds from low-frequency components, and auto-regressive moving average (ARMA) models are employed to simulate random fluctuations from the high-frequency components. We evaluate the framework on a traffic dataset collected in Chengdu, China, and experimental results demonstrate that Wavelet-HST outperforms six state-of-art prediction methods by an improvement of 7.8% ∼10.5% in the root mean square error.

INDEX TERMS Traffic prediction, graph convolutional network (GCN), spatio-temporal modeling, higher-order connectivity patterns, wavelet transform, time-frequency properties.

I. INTRODUCTION

Predicting short-term traffic conditions in urban areas is of paramount importance for travel planning and traffic control. For instance, it enables travelers to foresee the potential congestions to make more appropriate travel routes, and then improve traffic capacity on a large-scale road network [1]. With the increase of urban road vehicles and the construction of intelligent transportation system (ITS) [2], rich trajectory and flow data are collected by variety of sensors (e.g. GPS detectors equipped on floating vehicles and loop detectors fixed on roads), which provide information about the infrastructure and data environment enabling traffic prediction [3]. Furthermore, the emergence of big data technology helps us understand traffic dynamics in detail, and thus makes traffic prediction become a hotpot in transportation research. In this paper, we study the problem of traffic speed forecasting on road networks using historical traffic speed observations. Such a network-wide prediction task is challenging due to the complex spatial and temporal dependencies between road segments.

In a road network, the traffic condition on one road segment is affected by its locally adjacent road segments, and closer segments have a stronger influence on a given segment than more distant segments. This traffic interaction is not only based on the spatial proximity in Euclidean space, but also strongly restricted by the topologic structure of road networks. Therefore, an accurate representation of the road network is the key to model spatial correlations of traffic speeds. In previous studies, researchers rearranged road segments as

The associate editor coordinating the review of this article and approving it for publication was Zhengbing He.

one-dimensional sequences [4] or transformed road networks as two-dimensional images [5], and then used convolutional neural network (CNN) to capture the geographic similarity in local areas. Although they achieved promising performance, the predictions were limited to ring roads only or to areas in a cell-based tessellation. Moreover, as some road segments with close locations but long road network distances show very different behaviors (e.g. A two-lane road with opposite directions) [6], the spatial dependencies captured by traditional CNN-based methods are not consistent with real traffic dynamics. To address this issue, researchers further introduced graphs to represent road networks and employed graph convolutional networks (GCNs) to extract the spatial correlations among adjacent road segments [7]. These methods account for the topological relations of the road network, but they are still generalized measurements of spatial proximity in network space (i.e. topological adjacency). Thus, the rich structural and connectivity information inherent in road network data have not yet been fully incorporated in traffic prediction models.

As a type of complex networks, road networks show many typical structural characteristics, among which the most obvious one is the local connectivity patterns. Compared with the individual link between two adjacent road segments, the combination of multiple links in local areas has more significant influence on the diffusion of traffic flow, which is termed higher-order connectivity patterns [8]. Figure 1 illustrates an example. The small road network consists of fourteen road segments and the combination of road links in local areas forms three types of higher-order connectivity patterns: two-hop, converging and diverging structures, as shown in Figure 1(f). In each structure, the traffic flow follows particular propagation paths and generates corresponding spatial correlations. As one road segment often participates in multiple local structures, the traffic state of a road is affected by different higher-order connectivity patterns. For instance, Figures 1(a) through 1(e) illustrate different higher-order connectivity patterns that affect road 4. The diffusion process of traffic flow on road 4 present a complex spatial dependency in the local areas given the integrated impact of these structures. The congestion on road 7 may cause most of the traffic flow on road 4 to move to road 5, while the congestion on road 12 will trigger the flow of vehicles on route 3 to shift onto route 4. Therefore, traffic in a road network forms a higher-order type of spatial correlation, which is not based on adjacent road segments, but on adjacent local structures. A road network represented by these higher-order structures can describe the traffic interaction in more details than general graphs.

At the same time, for each road segment, traffic speed time-series data display multiple fluctuations at different frequencies. For instance, the smooth and periodic profile of speed sequences show a low-frequency changing trend, while the real-time traffic dynamics exhibits high-frequency random shakings. These mixed time-frequency properties increase the difficulties of capturing the spatio-temporal dependencies of traffic speeds. Some studies therefore took advantage of the multi-resolution analysis capability of wavelet decomposition, to obtain separate sub-time series with frequencies ranked from high to low [9], [10]. They employed different prediction models for different frequency sequences or simply remove high-frequency parts as noise. Despite high prediction accuracy, these methods rarely considered the correlation of multiple low-frequency sequences, and thus cannot capture more spatial details from stationary evolution patterns. Furthermore, the low-frequency components decomposed from raw traffic data show multi-scale time dependency. In the long term, the fluctuation of low-frequency sequences has a repetitive daily period, and in the short-term, the signal has similarity and continuity in the moments before and after the current moment, which can be referred to as *daily period* dependency and *recent trend* dependency, respectively [11]. These complex multi-frequency properties and multi-scale characteristics of traffic time series make it more challenging to learn temporal dependencies of traffic speeds.

To tackle these challenges, we consider a road network as a directed graph and propose a hybrid prediction framework that effectively learns the spatio-temporal dependencies of traffic speeds. The main contributions of this paper can be summarized as follows:

1) We employ motifs (local sub-graph structures [8]) to define the higher-order connectivity patterns presented in road networks. Meanwhile, we integrated these road motifs with the graph convolution neural network (GCN) to extract higher-order spatial correlations of traffic speeds, which is different from existing deep learning methods that only consider the geographical proximity in Euclidean space or the low-order topological adjacency on general graphs. To the best of our knowledge, this is the first work that applies motifs into traffic prediction research.

2) We propose a novel hybrid framework for traffic speed prediction, namely wavelet-based higher-order spatio-temporal (Wavelet-HST) method. It decomposes raw traffic speed data into several time-frequency sequences based on discrete wavelet transform (DWT), and then conduct multi-resolution analysis. Specifically, we employ ARMA to model the unstable randomness fluctuations from high-frequency components, and propose a spatio-temporal neural network – Motif-GCRNN to capture the stationary evolution patterns from low-frequency components, in which two LSTMs are used to learn *recent trend* and *daily period* temporal dependencies respectively. Through multi-level description, the framework provides a comprehensive picture of traffic dynamics.

3) Our experimental results on a real-world road network demonstrate that the proposed hybrid framework Wavelet-HST outperforms existing baseline methods.

The rest of this paper is organized as follows. Section II gives an overview of the related work. Section III introduces some preliminary definitions that will be applied in this paper Section IV presents the proposed framework Wavelet-HST

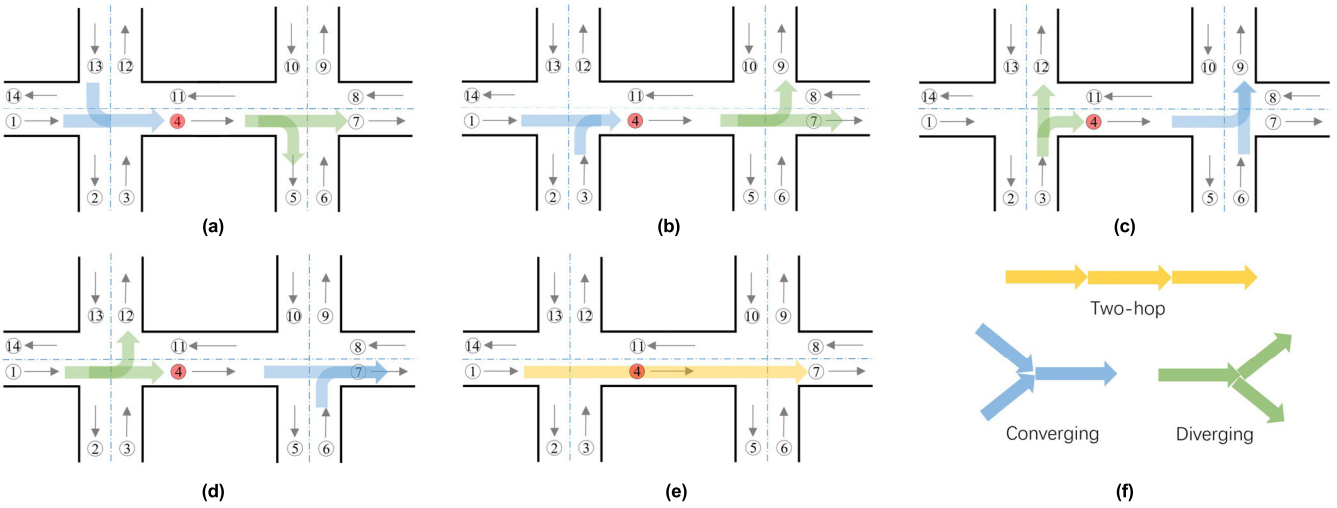


FIGURE 1. Higher-order connectivity patterns in a road network.

in detail and the experiments are carried out in Section V. Finally, Section VI summarizes the research of this paper and provides the future direction.

II. RELATED WORK

A. SHORT-TERM TRAFFIC PREDICTION

Existing research on short-term traffic prediction can be divided into two categories: time series analysis based on statistical approach and data-driven methods based on machine learning. Statistical approaches such as Kalman Filtering [12] and Auto-Regressive Integrated Moving Average (ARIMA) [13] achieve promising results but rely on stationary assumptions. With the increase of traffic datasets and the development of computational power, machine learning approaches, such as K-Nearest Neighbors (KNN) [14], Support Vector Regression (SVR) [15] can capture complex non-linear relations in traffic data but still fail to model spatio-temporal dependencies effectively.

Fortunately, numerous attempts were made to capture spatial and temporal dependencies simultaneously. Zheng *et al*. proposed a feature selection based method to predict the urban traffic speed with the consideration of spatiotemporal traffic pattern [16]. Most recently, deep learning methods have been widely applied in traffic prediction domain. Zhao *et al.* regarded traffic data as time series and used Long Short-Term Memory network (LSTM) to capture the temporal dependencies of traffic speeds [17], [18], [19]. However, these time-series prediction methods are only performed on a single road segment. When they are applied to the whole road network, the computational efficiency will be reduced from multiple prediction tasks. Several works found that learning multiple tasks jointly can improve the prediction performance compared with learning them individually [20], [21]. Ma *et al.* employed Convolutional Neural Network (CNN) to extract the spatial dependencies in traffic flow but the road networks studied in these research were expressed in the Euclidean space (e.g., 2D images) [1], [4], [5], [22]. The local spatial correlations learnt therefore only reflect unstructured spatial proximity, rather than structured adjacency relations of road segments. Yu *et al.* further introduced graphs to represent the topological structures of road networks [7], [23] and investigated graph convolution [24]. Lin *et al*. proposed a graph convolutional neural network with data-driven graph filter (GCNN-DDGF) that can learn hidden heterogeneous pairwise spatial correlations [25]. However, these graphs were set undirected and thus cannot describe the directed propagation of traffic flow along road networks. To this end, Li proposed the diffusion convolution operation on directed graphs to capture the spatial correlations of traffic flow [6], but this diffusion process was only defined by a state transition matrix at the level of individual nodes and edges. Thus the spatial information obtained by diffusion convolution only reflects low-order connectivity patterns of road segments.

B. WAVELET-BASED HYBRID PREDICTION

Previous studies show that the wavelet-based methods can effectively improve the prediction accuracy from noisy and unstable traffic data series. Some research employed wavelet decomposition to eliminate the noise of raw traffic data and then fed the remainder stationary sequences into statistical models or neural networks for traffic prediction [9], [26]. As a consequence, their prediction results lack some details compared with the ground truth data since the random information generated from context factors is removed along with noise in the denoising process. Based on the multi-resolution analysis (MRA) capability of wavelet decomposition, some studies further developed hybrid forecasting approaches that maintains different time-frequency characteristics of traffic series. For instance, Sun *et al*. proposed a hybrid method to forecast different kinds of passenger flows in the subway system [27]. It employs discrete wavelet transform (DWT) [28] to decompose the passenger flow data into several high-frequency and low-frequency series, and then predicts each

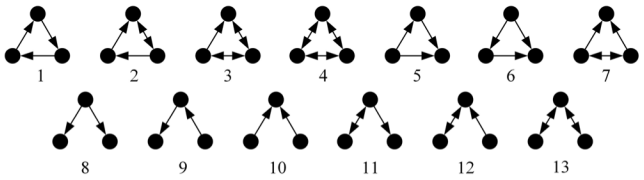


FIGURE 2. Thirteen triangular motifs.

frequency series with support vector machines (SVM). Their predicted flows therefore show different degree of random fluctuations. However, these hybrid methods only take individual time series data as model inputs, ignoring the correlation between different sequences, and thus cannot capture the spatio-temporal dependences implicit in traffic data.

III. PRELIMINARIES

In this section, we briefly introduce three basic concepts and technologies applied in our prediction framework, as a preparation for section IV.

A. HIGHER-ORDER NETWORK STRUCTURES

Networks are a fundamental tool for understanding and modeling complex systems in sciences, which can be expressed as graphs – combination of nodes and edges. However, many networks in real world have been shown to include higher-order connectivity patterns that control and mediate the behavior of complex systems at the level of small subgraphs [8]. Since these higher-order connectivity patterns appear frequently in a network, they can be treated as a new type of basic building blocks for complex networks. Milo et al. defined network motifs to represent these local meaningful sub-structures [29]. Figure 2 presents the commonly used triangular motifs, in which different motifs convey different interactive patterns. In current research, these motifs are often used in social networks for clustering analysis or in biological neural networks for information transmission pattern detection [30], [31], [32], but their applications in road networks for transportation research have not been touched yet.

B. GRAPH CONVOLUTIONAL NETWORK

In deep learning models, convolution can extract spatial features and reduce training parameters efficiently by an operation of local weighted average. However, the traditional convolution filter in CNNs is defined for regular grids in Euclidean space, which is not applicable to general graphs. Currently, there are two basic approaches to implement graph convolution: the spatial approach and the spectral approach. Spatial approaches rearrange the vertices into the form of a grid, processed by regular convolutional operations. This provides accurate filter localization but faces the challenge of matching local neighborhoods [33]. Spectral approaches define the convolution operation via a graph Fourier transform in the spectral domain. Bruna *et al.* proposed a graph convolution neural network based on the spectrum of the graph Laplacian, however their approach has high computational complexity due to the cost of computing a graph Fourier transform [34]. Defferrard *et al.* then designed fast localized convolutional filters by parametrizing the spectral filters as Chebyshev polynomials (referred to as ChebNet), which offers linear computational complexity [24]. A ChebNet filter can be written as:

$$g_\theta(L) = \sum_{k=0}^{K} \theta_k T_k(\tilde{L}) \tag{1}$$

where $\tilde{L} = 2L/\lambda_{\max} - I_n$ is the rescaled Laplacian matrix with maximum Eigenvalue $\lambda_{\max}$, the parameter $\theta \in R^K$ is a vector of Chebyshev coefficients, and $T_k\left(\tilde{L}\right) \in R^{n\times n}$ is the Chebyshev polynomial of order k. K is the number of successive filtering operations or convolutional layers and K localized convolution filters effectively exploit the information from the K-1-order neighborhood of a node.

C. DISCTRETE WAVELET TRANSFORM

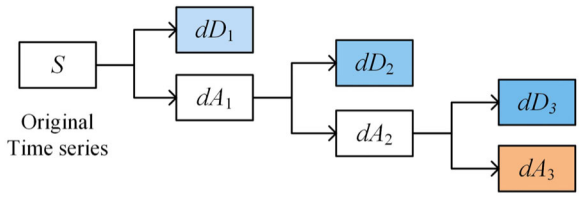
Wavelet transform (WT) is often used to extract information in the analysis of non-stationary data. This transform provides localization properties in both time and frequency domains [35]. In our approach, discrete wavelet transform (DWT) is employed to decompose raw traffic speed data into several frequency sequences, so as to conduct multi-resolution analysis (MRA). An effective way to perform DWT is through the Mallat algorithm [36] that passes data through a series of low-pass and high-pass filters:

$$dA = \sum_{k=-\infty}^{\infty} S[k]\varphi_l[2n-k] \tag{2}$$

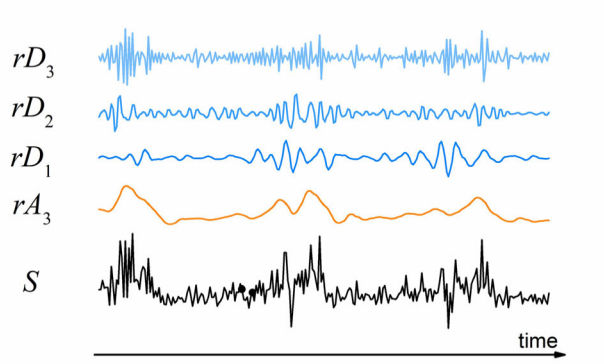
$$dD = \sum_{k=-\infty}^{\infty} S[k]\varphi_h[2n-k] \tag{3}$$

where S is the original signal, φ_l is low-pass filter and φ_h is high-pass filter, and dA and dD are the outputs of low-pass and high-pass filters, called approximation and detail coefficients, respectively.

Figure 3(a) illustrates the process of Mallat algorithm with three-level decompositions. The original time series data S is firstly passed through both low-pass and high-pass filters to obtain the approximation and detail coefficients (i.e., dA_1 and dD_1) at the first level. The obtained approximation coefficient dA_1 is then passed through both filters again to obtain two coefficients, dA_2 and dD_2, at the second level. This process is repeated until the specified level has been reached. The last approximation coefficient dA_3 and all detail coefficients (dD_1, dD_2 and dD_3) are retained after the decomposition process and wavelet reconstruction is performed on each of these coefficients. Specifically, when reconstructing time series from one coefficient, all coefficients except this one are set as zero values, and fed into the reconstruction algorithm (i.e., inverse discrete wavelet transform, IDWT). As shown in Figure 3(b), the low-frequency sequences rA_3 is



(a) three-level decomposition process



(b) reconstruction results

FIGURE 3. Discrete Wavelet Transform.

reconstructed using the approximation coefficient dA_3, which exhibits a smooth trend in fluctuation, and high-frequency sequences with different frequency sub-bands are obtained from reconstruction results of the detail coefficients, which show stochastic change patterns.

IV. METHEDOLOGY

A. PROBLEM FORMULATION

We use an directed graph $G = (V, E, A)$ with N nodes to describe a road network, where nodes $v_i \in V$ denote road segments, edges $(v_i, v_j) \in E$ indicates the directed connection from node i to node j and $A \in R^{N\times N}$ $\left(A^T \neq A\right)$ represents the adjacency matrix. When there is an edge from node i to node j (i.e. there is an adjacent link from road segment i to road segment j), $A_{ij} = 1$, otherwise, $A_{ij} = 0$ $(A_{ii} = 0)$. In the real world, as traffic speeds in two opposite directions of the same road could differ greatly at the same time interval, they must be predicted separately. Accordingly, in this study, any road in the real world with two lanes going in opposite directions is considered as two uni-directional road segments, that is, two nodes in a graph.

At the time interval t, we use $x_t^{v_i}$ to represent the average traffic speed on the road segment v_i and a speed vector $X_t = \left[x_t^{v_0}, x_t^{y_1}, \ldots, x_t^{y_{N-1}}\right]$ to represent the speed information of all road segments. Given the historical speed observations until time interval t on road segments in a road network, the task aims to predict traffic speeds at time interval $t + 1$, which can be formulated as:

$$X_{t+1} = F\left(\left[X_{t-T+1}, \ldots, X_{t-2}, X_{t-1}, X_t\right], G(V, E, A)\right) \quad (4)$$

where T is the length of historical observations, G is the directed Graph built with the road network, and F is the prediction function that must be learned.

B. FRAMEWORK OVERVIEW

Figure 4 presents the framework of Wavelet-HST. Original traffic speed data are processed by discrete wavelet transform (DWT) to obtain multiple frequency components for each road segment. A spatial-temporal prediction network – Motif-GCRNN is then proposed to capture the stationary spatio-temporal patterns from low-frequency components for all road segments, and auto-regressive moving average (ARMA) models are employed to simulate random shakings occurred in the high frequency components. Outputs of different frequency components are summarized by road segment to obtain the final prediction results.

Figure 5 illustrates the architecture of Motif-GCRNN. As shown in the middle part of Figure 5, we construct motif-based graph convolutional (MGC) layers that capture spatial correlations among road segments and use a recurrent layer with two LSTMs that learn temporal dependencies of traffic dynamics. At each time interval, the data is a speed vector of all road segments, represented by a simplified notation of a graph as shown in the yellow and blue squares, in Figure 5. We select historical speeds over the past few time intervals and speeds at the same time over the past few days as input to predict the speed at the next time interval. These two parts of input are fed into MGC layers respectively, and then spatial features extracted are reshaped to feed into LSTM for learning short-term and periodic patterns respectively (i.e., *recent trend* and *daily period*). These two types of temporal features are concatenated and fed into a fully connected layer for the prediction outputs.

C. MOTIF-BASED GRAPH CONVOLUTIONAL LAYER

In a road network, the speed prediction for one road must account for the traffic conditions on adjacent roads. This spatial interaction is based on the topological structure of the road network, but more prominently affected by the higher-order connectivity patterns of local adjacent road segments. In order to capture these particular spatial dependencies in traffic situation, we need to design a more meaningful representation form of road networks which can integrate higher-order connectivity patterns into general graphs. Motivated by the work in [8], [30], [31] and [32], we introduced motifs to describe the local higher-order connectivity patterns and defined a motif-based adjacent matrix to represent the road network. Furthermore, on the basis of ChebNet [24], we constructed motif-based graph convolution that extracts higher-order spatial correlations of traffic speeds efficiently.

1) MOTIFS IN ROAD NETWORKS

Since the structure formed by three objects is the most common and simplest basic structure in real world, and the combination of a road segment and its first-order adjacent road segments (i.e. upstream and downstream) also reflects

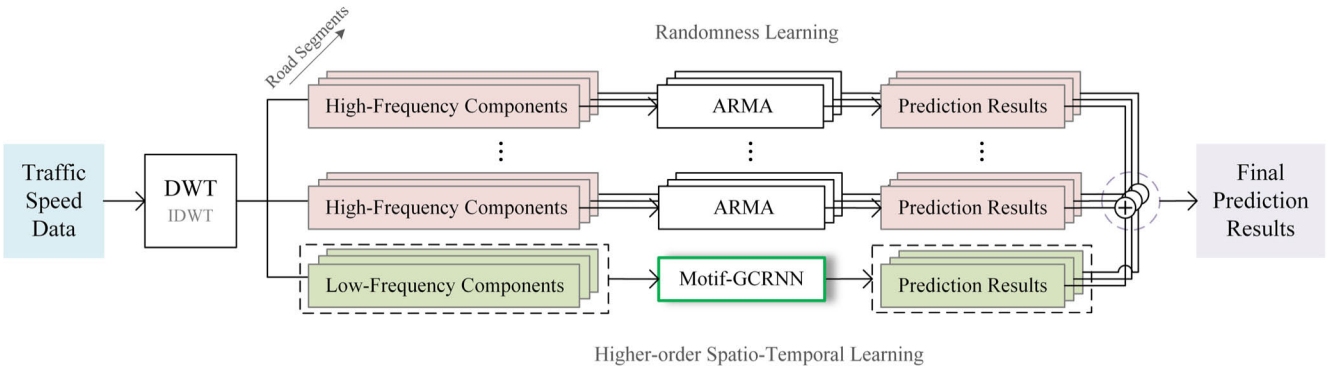


FIGURE 4. The framework of Wavelet-HST (Wavelet-based Higher-order Spatial-Temporal model). DWT: Discrete Wavelet Transform; IDWT: Inverse Discrete Wavelet Transform; ARMA: Auto-Regressive Moving Average. Motif-GCRNN: Motif-based Graph Convolutional Recurrent Neural Network.

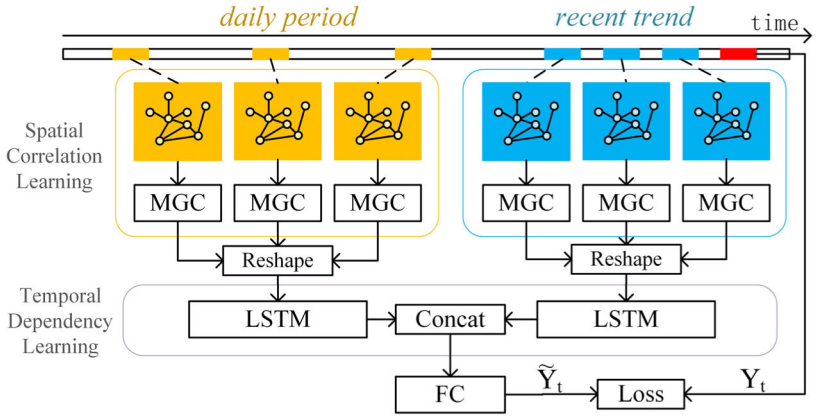


FIGURE 5. The architecture of Motif-GCRNN. MGC: Motif-based Graph Convolution; LSTM: Long Short-Term Memory; FC: Fully Connected.

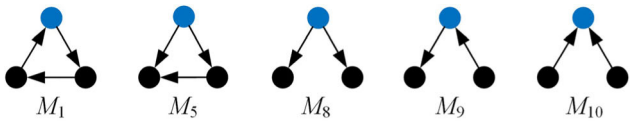


FIGURE 6. Motifs in Road Networks.

triangular relationship, thus in our model, triangular motifs are selected as fundamental units building blocks to construct a new representation of road networks. More complex road structures (e.g. mixed adjacency structure of second- and third- order) can be described by combination of these triangular motifs.

As shown in Figure 6, five motifs with only uni-directional edges among thirteen types of triangular motifs, were selected to represent possible local higher-order structures in urban road networks. The blue nodes in these motifs are anchor nodes that are regarded as the research target, and nodes in black represent the nodes that are adjacent to the target node. From the perspective of a road network, anchor nodes represent the target road segments that have influence on or are influenced by the adjacent road segments in local area. Based on the analysis in the Introduction section, M_8, M_9, M_{10} in Figure 6 can represent the diverging, two-hop, converging structure respectively, while M_1 and M_5 may reflect a ring structure and a detour structure occurred rarely in road networks.

2) MOTIF-BASED ADJACENCY MATRIX

Taking advantage of these sub-graphs, we constructed a motif-based adjacent matrix W_M that stores the higher-order spatial information of a road network. As defined before, we use an unweighted directed graph to represent this road network. For each edge $(v_i,v_j) \in E$, let $w_{k,ij}$ denote the number of times that the edge (v_i,v_j) participates in M_k ($k \in \{1, 5, 8, 9, 10\}$). The motif-based adjacency matrix W_M is then defined as:

$$(W_M)_{ij} = \sum_k w_{k,ij} \tag{5}$$

Each weight in matrix W_M is a statistical value that records the degree of involvement of an edge in these five motifs. For a road network, this weight describes the importance of an adjacent link between two road segments in local areas under the integrated impact of multiple higher-order connectivity patterns. Taking the road network in Figure 7(a) for example, the edge between road 1 and road 2 appears in the four motifs in local area, including the diverging structure (i.e. M_8) of road 1, road 2 and road 12, the diverging structure (i.e. M_8) of

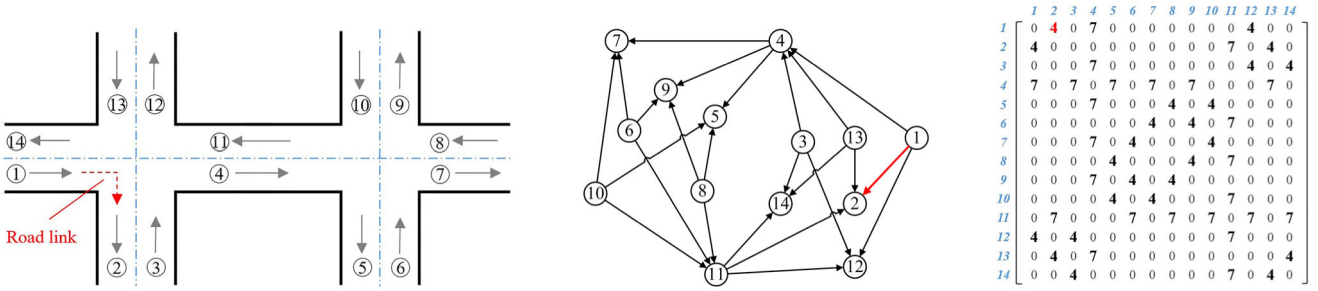


FIGURE 7. Expression of higher-order spatial correlations in Motif-based Adjacency Matrix. The weight in Motif-based Adjacency Matrix describes the importance of an adjacent link between two road segments in local areas under the integrated impact of multiple higher-order connectivity patterns.

road 1, road 2 and road 4, the converging structure (i.e. M_9) of road 1, road 2 and road 13, and the converging structure (i.e. M_9) of road 1, road 2 and road 11. Thus the number of times that the edge between road 1 and road 2 participates in the five motifs is four. In other words, the weight between road 1 and road 2 in the motif-based adjacency matrix is four.

3) MOTIF-BASED GRAPH CONVOLUTION

Through the weights derived from the adjacency matrix, a local convolution filter can effectively integrate the effects of diverse connectivity patterns between adjacent roads, to forecast traffic speeds on the central road. Based on the convolution operation as shown in (1), we define the motif Laplacian associated with the adjacency matrix W_M as $\tilde{\Delta} = I - \tilde{D}^{-1/2} W_M \tilde{D}^{-1/2}$ ($\tilde{D}$ is the degree matrix of W_M), which acts anisotropically in a preferred direction along structures associated with the motifs. We use this Laplacian to calculate the motif-based convolution on the directed road graphs $G(V, E, A)$ with the low frequency components X^{low} of traffic speeds at time interval t:

$$h_t = \mathbf{\Theta}\left(X_t^{low}, G\right) = \sigma\left(\sum_{k=0}^{K} \theta_k T_k\left(\tilde{\Delta}\right) X_t^{low}\right) \tag{6}$$

where $\mathbf{\Theta}$ is the graph convolution filter, h_t is the output of graph convolution, σ is the activation function (e.g., ReLU), $\theta \in R^{k \times n}$ is the trainable parameter, T_k is the Chebyshev polynomial of order k, and K is the size of filters' reception fields. Although road motifs only contain first-order adjacency information, the spatial correlation of multi-order adjacency can be effectively extracted by the motif-based graph convolution with an appropriate size of filters' reception fields K.

D. RECURRENT LAYER

The dynamic trend over recent time intervals and the periodic repetitions at a daily scale in traffic speed data show strong regularity. Taking full advantage of these temporal patterns can help improve prediction performance. As Long Short-Term Memory (LSTM) [37] is a powerful deep learning method capable of learning long-term temporal patterns of a time series, we use a recurrent layer with two LSTM to explore the short-term and periodic dependency of traffic speeds, what we term the *recent trend* and *daily period*, respectively. Before using the recurrent layer, we reshape the spatial features extracted by MGC layers in the past m time intervals and the same time interval of past n days into the input form of LSTM separately. In LSTM, the gates structure and the hidden state are unchanged, only the input is replaced by these spatial features. Assume the predicted speed is at the t time interval of the d day, the *recent trend* and *daily period* features are defined as:

$$Y_{\text{rend}} = LSTM\left(\left[h_{\text{t-m}}^{(d)}, \ldots, h_{\text{t-2}}^{(d)}, h_{t-1}^{(d)}\right]\right) \tag{7}$$

$$Y_{\text{period}} = LSTM\left(\left[h_t^{(d-n)}, \ldots, h_t^{(d-2)}, h_t^{(d-1)}\right]\right) \tag{8}$$

where $h_t^{(d)}$ is the output of graph convolution at the t time interval in the d day, Y_{trend} represents the *recent trend* spatio-temporal features extracted from LSTM, and Y_{period} represents the *daily period* spatio-temporal features.

To fuse these two types of spatio-temporal dependencies for future traffic prediction, we concatenate them as Y_{concat}.

E. REGRESSION LAYER

We use a fully connected layer to learn the comprehensive impact affected by *recent trend* and *daily period*:

$$\tilde{Y}_t = tanh\left(W_{FC} \circ Y_{concat} + b_{FC}\right) \tag{9}$$

where Y_{concat} is the concatenation of *recent trend* and *daily period* features, W_{FC} indicates the learnable parameters, $\circ$ is the element-wise multiplication operator, b_{FC} represents the bias in the fully connected layer, tanh is the activation function of hyperbolic tangent that ensures the output values are between -1 and 1, and $\tilde{Y}_t$ is the prediction outputs of Motif-GCRNN, corresponding to the predicted low-frequency component of traffic speeds at time interval t.

The Motif-GCRNN is trained by minimizing mean squared error between predicted low-frequency vectors $\tilde{Y}_t$ and true low-frequency vectors Y_t, the loss function is defined as:

$$\ell(\theta) = \sum_t \left\| Y_t - \tilde{Y}_t \right\|^2 \tag{10}$$

where θ are all learnable parameters in Motif-GCRNN.

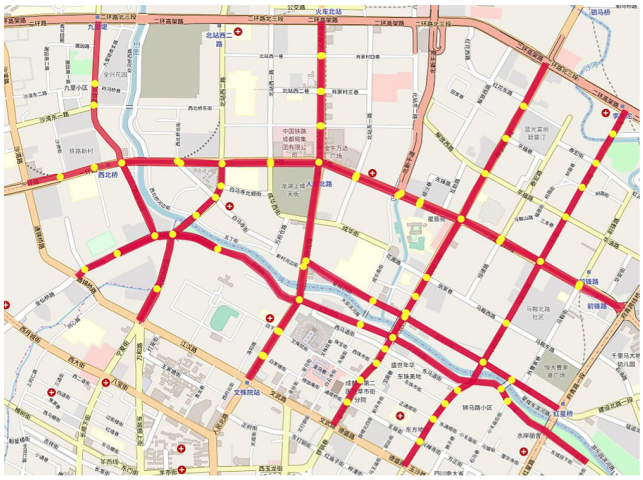


FIGURE 8. Road network for testing.

F. ARMA MODEL

Through discrete wavelet transform, the high-frequency components decomposed from original series show a strong random pattern. Since Auto-Regressive Moving Average (ARMA) is one of the most common models used in stationary stochastic process analysis in the time series domain [38], we employ it to train and predict the high-frequency components of traffic speeds for simulating the random shakings occurred in raw traffic data. The outputs will be added together to form the final prediction results. ARMA consists of two polynomials, one for the auto-regression (AR) and the second for the moving average (MA). The calculation process can be represented as:

$$\tilde{y}_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i}^{high} + \sum_{i=1}^{q} \lambda_i \varepsilon_{t-i} \tag{11}$$

where c is a constant, ε is the random error and distributed as a Gaussian white noise, φ, λ are parameters of AR and MA part, p and q are integers showing the orders of AR and MA part respectively, and $\tilde{y}_t$ is the prediction outputs of ARMA, corresponding to the predicted high-frequency components of traffic speeds at time interval t.

V. EXPERIMENTS

A. DATASET DESCRIPTION

We verified the proposed framework using a taxi GPS dataset collected from 1st November 2016 to 30th November 2016 (30 days) in Chengdu, China. The trajectory of 43257 taxis were collected and the sample interval of GPS was three seconds. We calculated the travelling speed between two continuous GPS points of one car as a speed record and thus generated 63171201 speed records.

The highlighted roads in Figure 8 constitute the road network to be predicted, which consists of 156 uni-directional road segments and 210 directed edges of graph. The average length of the road segments was 300 meters and the average number of speed records for each road segment was 399818. All speed records during 30 days were aggregated into a 15 min interval, and the traffic speed of each road segment at each time interval was then obtained by averaging the speed records within this time interval for each road segment. Finally, we got 2880 speed samples on 156 road segments. In our experiment, data samples from the first 24 days are used as training data, and the remaining six days as test data.

TABLE 1. Performance comparison of different models.

Model	MAE	MAPE (%)	RMSE
ARMA	3.952	14.135	5.308
SVR	3.708	13.905	5.244
ST-CNN	3.624	13.471	5.211
LSTM	3.566	13.178	5.176
GCRNN	3.508	13.120	5.116
DCRNN	3.518	13.027	5.155
Wavelet-HST*	**3.287**	**12.112**	**4.700**

B. EXPERIMENTAL SETTINGS

1) BASELINES

We compare our framework with the following baseline methods: (1) Auto-Regressive Moving Average (ARMA); (2) Support Vector Regression (SVR); (3) Space-Time Convolutional Neural Network (ST-CNN) [4], which uses a speed matrix with x-axis of time and y-axis of roads to represent the spatio-temporal evolution of traffic speeds; (4) Long Short-Term Memory (LSTM) with two recurrent layers; (5) Graph Convolutional Recurrent Neural Network (GCRNN) [39], which integrates GCN and LSTM to predict structured sequence data. (6) Diffusion convolutional recurrent neural networks (DCRNN) [6] that integrates diffusion convolution and GRU units with an encoder-decoder architecture.

2) PARAMETER SETTINGS

In our approach, we used DB4 as mother wavelet for DWT and selected the decomposition level $j = 3$ referring to the work [26]. We then obtain three high-frequency components and one low-frequency sequence for training and prediction. For the low frequency components, we use the Min-Max normalization method to scale the input data into the range $[-1, 1]$ before the model training. After the model prediction, we re-scale the predicted value back to the normal values. In Motif-GCRNN, we use one convolutional layer with 32 filters and a max-pooling layer at a size of 2×2. The size of filter reception field K varies from one to five in motif-based convolutional layers, the number of *recent trend* intervals varies from one to eight (15min to 2 hours) and the size of *daily period* intervals varies from one to eight days in LSTM layers. The filter reception fields K was set to three, the number of *recent trend* intervals to two (half an hour) and *daily period* intervals were set to seven as default parameters. All neural network based approaches were implemented using Tensorflow, and trained using the SGD optimizer.

TABLE 2. Ablation study of Motif-based Graph Convolution.

Model	MAE	MAPE (%)	RMSE
GCRNN	3.508	13.120	5.116
Motif-GCRNN	**3.499**	**13.035**	**5.098**
DWT + GCRNN + ARMA	3.334	12.403	4.748
DWT + Motif-GCRNN + ARMA **(Wavelet-HST*)**	**3.287**	**12.112**	**4.700**

TABLE 3. Ablation study of Discrete Wavelet Transform (DWT).

Model	MAE	MAPE (%)	RMSE
ARMA	3.952	14.135	5.308
DWT + ARMA + ARMA	**3.412**	**12.766**	**4.915**
GCRNN	3.508	13.120	5.116
DWT + GCRNN + GCRNN	**3.506**	**13.092**	**5.103**
Motif-GCRNN	3.499	13.035	5.098
DWT + Motif-GCRNN + Motif-GCRNN	**3.491**	**12.878**	**5.063**

3) EVALUATION METRICS

We selected three commonly used metrics in traffic forecasting to evaluate the performance of the tested methods, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE). They are stated as equation (12), (13) and (14), respectively.

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|X_t - \left(\tilde{Y}_t + \sum_j \tilde{y}_t^j\right)\right| \tag{12}$$

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\frac{\left|X_t - \left(\tilde{Y}_t + \sum_j \tilde{y}_t^j\right)\right|}{X_t} \tag{13}$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(X_t - \left(\tilde{Y}_t + \sum_j \tilde{y}_t^j\right)\right)^2} \tag{14}$$

where T is the length of traffic time series, X_t is the observed traffic speed at time interval t, $\tilde{Y}_t$ is the predicted low-frequency component of traffic speeds at time interval t and $\tilde{y}_t^j$ is the predicted j-th high-frequency component of traffic speeds at time interval t.

C. EXPERIMENTAL RESULTS

1) COMPARISON BETWEEN DIFFERENT MODELS

Table 1 illustrates the performance results of our hybrid method Wavelet-HST and six other baselines.

We analyzed the results in Table 1 and found that overall, our proposed hybrid prediction framework Wavelet-HST significantly improves performance with the lowest test error (RMSE = 4.700) among the six other benchmarks tested, according to three metrics. These results verify the effectiveness and superiority of our method for short-term traffic forecasting on a road network. Meanwhile, RNN-based methods, including LSTM, GCRNN, DCRNN and Wavelet-HST, generally outperform other baselines, which emphasizes the importance of modeling the temporal dependency. Furthermore, the comparative results of convolution based methods, including ST-CNN, GCRNN, DCRNN and Wavelet-HST prove that the effectiveness of spatial correlation modeling is related to the representation of road network. Compared with ST-CNN which rearranges road networks as one-dimensional sequences, GCRNN expresses road networks as undirected graphs and thus captures more accurate topological correlations with higher prediction accuracy. DCRNN uses directed graphs to model the real diffusion process of traffic flow and further improves the prediction performance. In contrast with these three models, Wavelet-HST innovatively employs motifs to construct the road network and achieves the highest accuracy for all three metrics, indicating that considering higher-order connectivity patterns is significant for spatial dependencies learning.

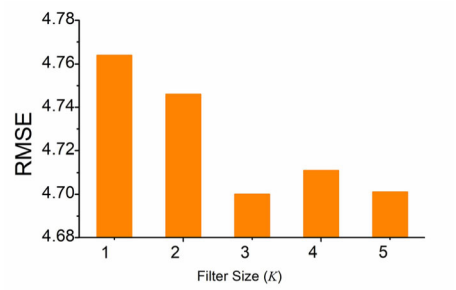
2) ABLATION STUDIES

In order to test the effects of each component in our prediction framework, we conduct three groups of ablation studies as shown in Table 2, Table 3 and Table 4. We select some reference models and set variants of Wavelet-HST with different components. For example, DWT + GCRNN + ARMA represents a DWT-based method that uses GCRNN to predict low frequency components and ARMA to predict high frequency components. Based on this type of model shortened form, DWT + ARMA + ARMA represents a classic statistical method that uses ARMA to predict both low and high-frequency sequences, and DWT + Motif-GCRNN + ARMA stands for our proposed framework Wavelet-HST itself.

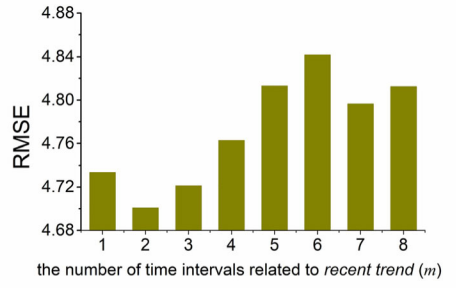
As shown in Table 2, Motif-GCRNN achieves higher accuracy than GCRNN for all three metrics, indicating that the introduction of motifs in road networks is effective. The proposed motif-based graph convolution extracts more

TABLE 4. The model comparison between low-high frequency components.

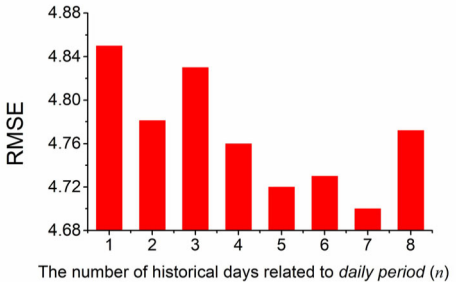
Model	MAE	MAPE (%)	RMSE
DWT + ARMA + Motif-GCRNN	3.522	13.145	5.148
DWT + Motif-GCRNN + Motif-GCRNN	3.491	12.878	5.063
DWT + ARMA + ARMA	3.412	12.766	4.915
DWT + Motif-GCRNN + ARMA **(Wavelet-HST*)**	**3.287**	**12.112**	**4.700**



(a) varying filter size *K*



(b) varying the number of time intervals related to *recent trend*



(c) varying the number of historical days related to *daily period*

FIGURE 9. Varying different parameters.

meaningful higher-order spatial correlations than general graph convolution and diffusion convolution. This indicates that the higher-order connectivity patterns presented in road networks have a more significant effect on the traffic interaction than pure topologically adjacent road segments. Furthermore, the test error of Wavelet-HST is lower than DWT + GCRNN + ARMA for three metrics, which also proves the validity of motif-based graph convolution block.

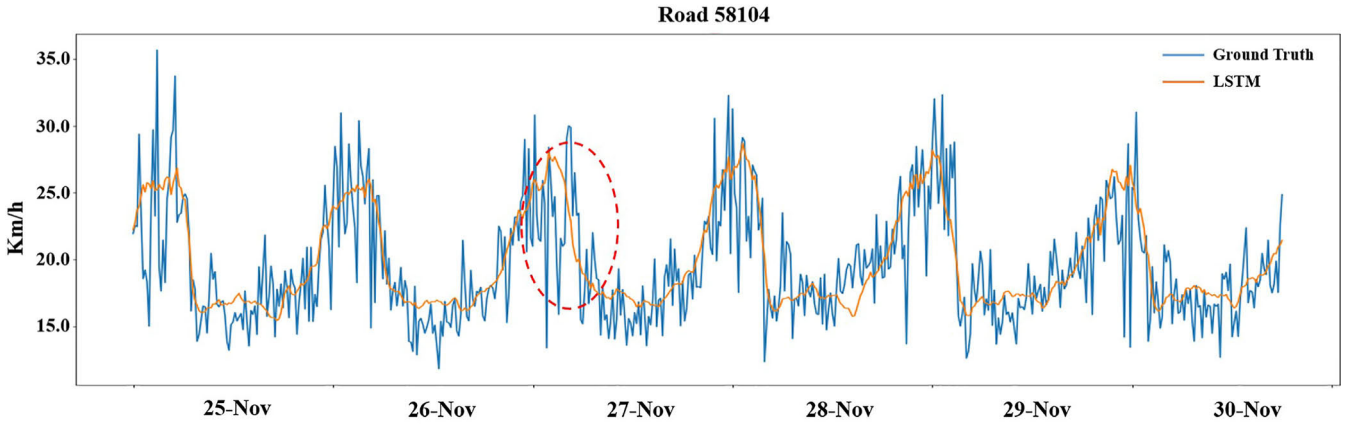
As shown in Table 3, we observed that the prediction accuracy of DWT-based models is higher than other methods without the pre-process of DWT, pointing out that multi-frequency properties of traffic speed data are beneficial for traffic prediction. Through decomposition and reconstruction with different frequency sub-bands, we can get more detailed information compared with the original traffic sequences. Intuitively, the variation characteristics of traffic time series at different frequencies and different scales are separated and emphasized, which makes it easier for models to learn the changing rules of a single time series

As shown in Table 4, compared with DWT + ARMA + ARMA and DWT + Motif-GCRNN + Motif-GCRNN, our proposed framework Wavelet-HST achieves lower test error, demonstrating that the frequency components with different changing characteristics should be predicted by different models. Furthermore, in contrast to DWT + ARMA + Motif-GCRNN, the proposed framework also shows better performance, indicating that our model selection strategy for different frequency components is correct. Intuitively, from the reconstruction results as shown in Figure 3(b), low-frequency sequences keep the fast-changing profile of traffic data series while high-frequency sequences present a large amount of random shakings. These inconsistent evolution mechanism is difficult to simulate by the same model. For the low-frequency components of traffic speeds, the complex spatio-temporal correlation is the dominant factor, so it is better to consider the interaction among adjacent road segments rather than learning the temporal patterns in a single road. For the high-frequency parts of traffic speeds, external context interference leads to the strong randomness in fluctuation, and thus stochastic process analysis based models are more suitable than complicated neural networks.

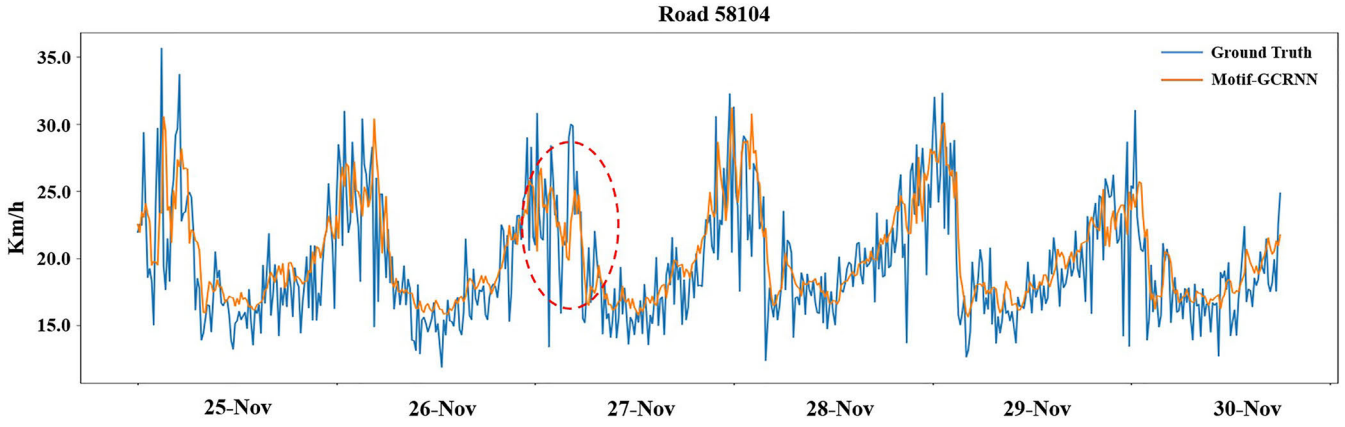
3) EFFECT OF DIFFERENT PARAMETERS

As shown in Figure 9, three parameters are selected to analyze the influences of spatio-temporal factors in Wavelet-HST on the prediction performance, including the size of filters' reception fields K, the number of *recent trend* time intervals m and the number of *daily period* time intervals n. The filter size K indicates the spatial range the filter can cover in convolution operation. For instance, when $K = 3$, adjacent road segments within three-order (including three-order) of the center road are fed into the convolution operation to capture the local spatial correlations of the center road. When we varied one parameter, other two parameters were set invariant with the default parameters that we defined before, in order to observe the individual impact of this factor.

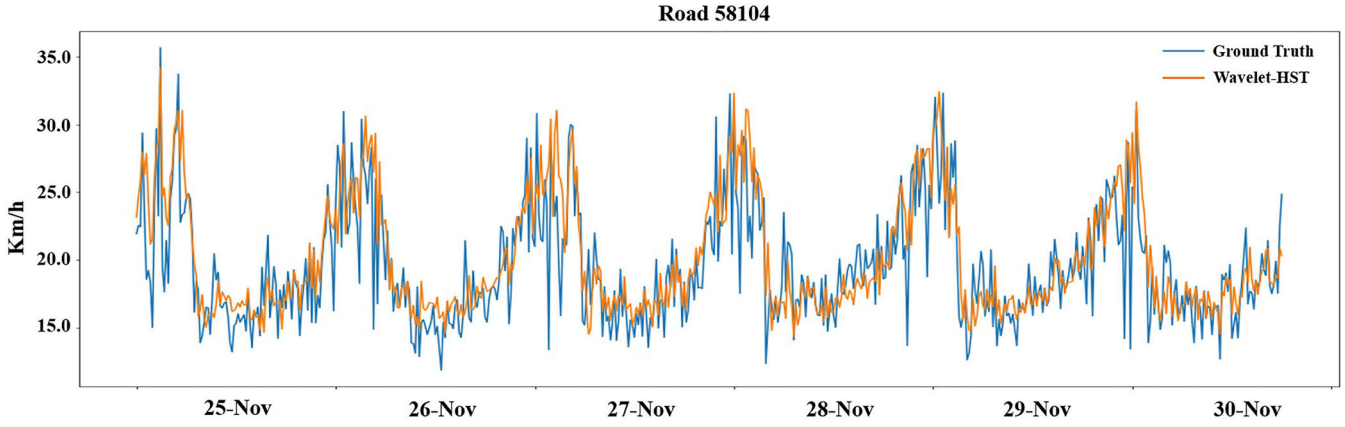
As shown in Figure 9(a), the RMSE of our method gradually decreases and then fluctuates slightly with an increase



(a) Prediction results on road 58104 by LSTM



(b) Prediction results on road 58104 by Motif-GCRNN



(c) Prediction results on road 58104 by Wavelet-HST

FIGURE 10. Traffic speed series forecasting visualization.

of K. The best prediction performance occurred when $K = 3$. Intuitively, this may be because a larger K enables the model to capture broad spatial dependency between the predicted segment and its adjacent road segments. However, the spatial dependency becomes weaker when the involved road segments within the reception field get farther from the predicted segment, indicating that the spatial information of distant road segments with higher adjacent order does not improve the prediction performance.

As shown in Figure 9(b), the model achieves the best performance when speed data of previous two time intervals (30 min) are fed into recurrent layers. Then, the RMSE goes

up with an increase in historical information. This result indicates that traffic dynamics have a strong short-term temporal dependency, and this dependency is concentrated over a short period in the past. Remote historical information therefore, is less valuable when capturing the temporal patterns of traffic speeds.

As shown in Figure 9(c), the prediction accuracy is improved by adding periodic data to the recurrent layer with the increase of historical days. The RMSE reaches a relatively low level when the length of *daily period* intervals is more than five. This result shows that traffic conditions at the same time interval during past days can be used to predict the possible traffic conditions in future days. The increase of test error occurred when the length of *daily period* intervals is eight; this may be due to irrelevant information accumulated from redundant historical data.

4) PREDICTION EFFECTIVENESS

To better understand the model, we visualize forecasting results of the last six days. Figure 10 shows the visualization of 15 min ahead forecasting on a random selected road segment ($K = 3$, *recent trend* interval $= 2$, and *daily period* interval $= 7$). We compare our framework with two typical methods: LSTM and Motif-GCRNN, in which LSTM is considered as a benchmark with no spatial correlation and Motif-GCRNN is chosen as a reference model with no wavelet-decomposition.

We have the following observations: (1) On a single road, the prediction results of LSTM (Figure 10(a)) clearly show the periodic fluctuation and the local upward or downward trend of traffic speed evolution, indicating that LSTM can effectively capture the temporal dependencies of traffic speeds. (2) At the peak of the curve inside the red dotted circle in Figure 10(a) and Figure 10(b), the prediction results of Motif-GCRNN are closer to the ground truth compared with LSTM. This suggests that the traffic status on a single road dose not just evolve over time, but also is affected by surrounding roads, and this spatial correlation can be successfully learnt by motif-based graph convolution. (3) As shown in Figure 10(c), Wavelet-HST captures more detailed information of traffic speeds than Motif-GCRNN on the whole time series, and the prediction results of LSTM are the most smoothing. This reflects Wavelet-HST is more likely to accurately predict abrupt changes in the traffic speed than other methods. Through exploring the multiple time-frequency properties of traffic sequences, the proposed DWT-based framework improves the prediction accuracy significantly.

VI. CONCLUSION AND FUTURE WORK

In this paper, we propose a novel hybrid framework Wavelet-HST for traffic speed forecasting. The wavelet transform is used to decompose different time-frequency components of raw traffic data so that the framework can learn both random and stationary patterns of traffic speeds. Specifically, a spatio-temporal network – Motif-GCRNN is proposed to predict the low frequency component and ARMA is employed to model high frequency components. In Motif-GCRNN, we introduce motifs to define the higher-order connectivity patterns of road segments and construct motif-based graph convolution that captures higher-order spatial correlations of traffic speeds. A recurrent layer with two LSTM is further used to learn short-term and periodic information, respectively. In practice, the proposed hybrid framework has shown to be effective in helping to improve the accuracy of urban traffic speed prediction.

In future work, we will move forward to explore the Dynamic Graph problem, in which the adjacent matrix of road networks is variant since the connection structure may be changed over time. And we will also apply the proposed hybrid framework to other spatial-temporal forecasting task.

REFERENCES

[1] X. Cao, Y. Zhong, Y. Zhou, J. Wang, C. Zhu, and W. Zhang, "Interactive temporal recurrent convolution network for traffic prediction in data centers," *IEEE Access*, vol. 6, pp. 5276–5289, 2018.

[2] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, Dec. 2011.

[3] U. Mori, A. Mendiburu, M. Álvarez, and J. A. Lozano, "A review of travel time estimation and forecasting for advanced traveller information systems," *Transportmetrica A, Transp. Sci.*, vol. 11, no. 2, pp. 119–157, 2015.

[4] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.

[5] H. Yu, Z. Wu, S. Wang, Y. Wang, and X. Ma, "Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks," *Sensors*, vol. 17, no. 7, p. 1501, 2017.

[6] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," in *Proc. 6th Int. Conf. Learn. Represent. (ICLR)*, 2018, pp. 1–16.

[7] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proc. 27th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2018, pp. 3634–3640.

[8] A. R. Benson, D. F. Gleich, and J. Leskovec, "Higher-order organization of complex networks," *Science*, vol. 353, no. 6295, pp. 163–166, 2016.

[9] Y. Xie, Y. Zhang, and Z. Ye, "Short-term traffic volume forecasting using Kalman filter with discrete wavelet decomposition," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 22, no. 5, pp. 326–334, 2007.

[10] J. Wang, Z. Wang, J. Li, and J. Wu, "Multilevel wavelet decomposition network for interpretable time series analysis," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2018, pp. 2437–2446.

[11] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 1655–1661.

[12] S. I.-J. Chien and C. M. Kuchipudi, "Dynamic travel time prediction with real-time and historic data," *J. Transp. Eng.*, vol. 129, no. 6, pp. 608–616, 2003.

[13] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 871–882, Jun. 2013.

[14] J. Rice and E. Van Zwet, "A simple and effective method for predicting travel times on freeways," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 3, pp. 200–207, Sep. 2004.

[15] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276–281, Dec. 2004.

[16] L. Zheng, C. Zhu, N. Zhu, T. He, N. Dong, and H. Huang, "Feature selection-based approach for urban short-term travel speed prediction," *IET Intell. Transp. Syst.*, vol. 12, pp. 474–484, Aug. 2018.

[17] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: A deep learning approach for short-term traffic forecast," *IET Intell. Transp. Syst.*, vol. 11, no. 2, pp. 68–75, Mar. 2017.
[18] D. Yanjie, L. Yisheng, and W. Fei-Yue, "Travel time prediction with LSTM neural network," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst.*, Nov. 2016, pp. 1053–1058.
[19] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-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.
[20] K. Zhang, Z. Liu, and L. Zheng, "Short-term prediction of passenger demand in multi-zone level: Temporal convolutional neural network with multi-task learning," *IEEE Trans. Intell. Transp. Syst.*, to be published. doi: 10.1109/TITS.2019.2909571.
[21] K. Zhang, L. Zheng, Z. Liu, and N. Jia, "A deep learning based multitask model for network-wide traffic speed predication," *Neurocomputing*, to be published. doi: 10.1016/j.neucom.2018.10.097.
[22] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.
[23] Z. Cui, K. Henrickson, R. Ke, and Y. Wang, "Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting," 2018, *arXiv:1802.07007*. [Online]. Available: https://arxiv.org/abs/1802.07007
[24] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 3844–3852.
[25] L. Lin, Z. He, and S. Peeta, "Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach," *Transp. Res. C, Emerg. Technol.*, vol. 97, pp. 258–276, Dec. 2018.
[26] H. Xiao, H. Sun, B. Ran, and Y. Oh, "Fuzzy-neural network traffic prediction framework with wavelet decomposition," *Transp. Res. Rec.*, vol. 1836, no. 1, pp. 16–20, 2003.
[27] 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.
[28] M. Shensa, "The discrete wavelet transform: Wedding the a trous and Mallat algorithms," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2464–2482, Oct. 1992.
[29] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: Simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, 2002.
[30] F. Monti, K. Otness, and M. M. Bronstein, "MOTIFNET: A motif-based graph convolutional network for directed graphs," in *Proc. IEEE Data Sci. Workshop (DSW)*, Jun. 2018, pp. 225–228.
[31] A. Sankar, X. Zhang, and K. C.-C. Chang, "Motif-based convolutional neural network on graphs," 2017, *arXiv:1711.05697*. [Online]. Available: https://arxiv.org/abs/1711.05697
[32] H. Peng, J. Li, Q. Gong, S. Wang, Y. Ning, and P. S. Yu, "Graph convolutional neural networks via motif-based attention," 2019, *arXiv:1811.08270*. [Online]. Available: https://arxiv.org/abs/1811.08270
[33] M. Niepert, M. Ahmed, and K. Kutzkov, "Learning convolutional neural networks for graphs," in *Proc. 29th Int. Conf. Mach. Learn.*, 2016, pp. 2014–2023.
[34] J. Bruna, W. Zaremba, A. Szlam, and Y. Lecun, "Spectral networks and locally connected networks on graphs," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2013, pp. 1–14.
[35] J. Morlet, G. Arens, E. Fourgeau, and D. Glard, "Wave propagation and sampling theory—Part I: Complex signal and scattering in multilayered media," *Geophysics*, vol. 47, no. 2, pp. 149–270, 1982.
[36] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
[37] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[38] M. S. Ahmed and A. R. Cook, "Analysis of freeway traffic time-series data by using Box-Jenkins techniques," *Transp. Res. Rec.*, vol. 773, no. 722, pp. 1–9, 1979.
[39] Y. Seo, M. Defferrard, P. Vandergheynst, and X. Bresson, "Structured sequence modeling with graph convolutional recurrent networks," in *Proc. Int. Conf. Neural Inf. Process.*, 2018, pp. 362–373.

NA ZHANG was born in 1996. She received the bachelor's degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2018. She is currently pursuing the master's degree in cartography and geography information system with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan. Her current research interests include deep learning, big data analytics, and intelligent transportation systems.

XUEFENG GUAN was born in 1980. He received the Ph.D. degree in cartography and geographical information engineering from Wuhan University, Wuhan, China, in 2011. He is currently an Associate Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan. His research interests include spatiotemporal data mining, distributed spatio-temporal database, and high-performance geo-computing.

JUN CAO was born in 1991. He received the bachelor's degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2014. He is currently pursuing the Ph.D. degree in cartography and geographical information engineering with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan. His current research interests include deep learning, graph convolutional networks, and traffic big data processing and analysis.

XINGLEI WANG was born in 1996. He received the bachelor's degree in geomatics engineering from Wuhan University, Wuhan, China, in 2018. He is currently pursuing the master's degree in cartography and geographical information engineering with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan. His current research interests include machine learning, data mining, and spatiotemporal forecasting.

HUAYI WU was born in 1966. He received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1999. He was a Postdoctoral Fellow with the Geospace Information and Communication Technology Laboratory, York University, Toronto, Canada, in 2002. He is currently a Professor and the Deputy Director with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan. His research interests include deep learning, data mining, and distributed computing. He is also a member and the Secretary General of the GIS Theory and Method Committee of China. He is also the Chairman of the IV/2 Working Group of the International Society for Photogrammetry and Remote Sensing.