Short-Term Traffic Forecasting: A Dynamic ST-KNN Model Considering Spatial Heterogeneity and Temporal Non-Stationarity

Shifen Cheng, Feng Lu

State Key Lab of Resources and Environment Information System

Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences

11A, Datun Road, Chaoyang District, Beijing 100101, P. R. China

{chengsf, luf}@lreis.ac.cn

ABSTRACT

Accurate and robust short-term traffic forecasting is a critical issue in intelligent transportation systems and real-time traffic related applications. Existing short-term traffic forecasting approaches are used to adopt global and static model structures and assume the traffic correlations between adjacent road segments within assigned time periods. Due to the inherent characteristics of spatial heterogeneity and temporal non-stationarity of city traffic, it is rather difficult for these approaches to obtain stable and satisfying results. To overcome the problems of static model structures and quantitatively unclear spatiotemporal dependency relationships, this paper proposes a dynamic spatiotemporal k-nearest neighbor model, named D-ST-KNN, for short-term traffic forecasting. It comprehensively considers the spatial heterogeneity and temporal non-stationarity of city traffic with dynamic spatial neighbors, time windows, spatiotemporal weights and other parameters. First, the sizes of spatial neighbors and the lengths of time windows for traffic influence are automatically determined by cross-correlation and autocorrelation functions, respectively. Second, dynamic spatiotemporal weights are introduced into the distance functions to optimize the search mechanism. Then, dynamic spatiotemporal parameters are established to adapt the continuous change in traffic conditions, including the dynamic number of candidate neighbors and dynamic weight allocation parameters. Finally, the D-ST-KNN model is evaluated using two vehicular speed datasets collected on expressways in California, U.S. and city roads in Beijing, China. Four traditional prediction models are compared with the D-ST-KNN model in terms of the forecasting accuracy and the generalization ability. The results demonstrate that the D-ST-KNN model outperforms existing models in all time periods, especially in the morning period and evening peak period. In addition, the generalization ability of the D-ST-KNN model is also proved.

1 INTRODUCTION

Short-term traffic forecasting, which has an important role in intelligent transportation systems, enables traffic managers to formulate reasonable and efficient strategies for alleviating traffic congestion and optimizing traffic assignments. Short-term traffic forecasting also enables the public to achieve accurate vehicular path planning [29][10].

In the past few decades, researchers have proposed several short-term traffic forecasting models that can be divided into two categories: parametric models and nonparametric models. A parametric model uses an explicit parametric function to quantify the relationship between historical traffic data and predicted traffic data. Considering the stochastic and nonlinear characteristics of traffic, constructing a mathematical model with high accuracy for characterizing traffic characteristics in practice is difficult [1]. Nonparametric models, such as data-driven methods, do not require a priori knowledge and explicit expression of mechanism; thus, they are more suitable for short-term traffic forecasting problems [22] [23] [7] [31].

As a typical nonparametric method, the k-nearest neighbors (KNN) model has received considerable attention. Many scholars have successfully applied the traditional KNN model to short-term traffic prediction [2][17][19][4][16][30][28]. Considering that the evolution of traffic is a spatiotemporal interaction process, traffic conditions of road segments are spatially and mutually affected [6]. Therefore, spatiotemporal relationships between multiple road segments in road networks are considered to improve traffic prediction [25][21][20]. Based on the traditional KNN model, [26] realized an enhanced model with the support of spatiotemporal information and argued that it achieves better performance than the model that employs only temporal information. [27] considered upstream and downstream traffic information and proposed a distributed architecture of a spatiotemporal-weighted KNN model for short-term traffic prediction. [3] employed a spatiotemporal state matrix instead of the traditional time series to describe the traffic state while using a Gaussian weight distance to select the nearest neighbor to improve the KNN model. However, the disadvantages of these ST-KNNs are that the spatiotemporal relation cannot be accurately quantified, which is primarily reflected in the modeling process, the size of the spatial dimension m and the length of time window n of the state space cannot be automatically determined, and some values are artificially set. For example, for m=3, three adjacent road segments are selected; for n=2, the historical data of the first two time steps of the current time step are used to construct samples. When the time series problem is transformed into a supervised machine learning problem, the values of m and n determine the number of selected features. Therefore, manually engineered features can easily cause dimensional disaster prevent the guarantee of the prediction accuracy of the model [15].

The prediction model is usually static, thus, it cannot describe the characteristics of the dynamic change in traffic, which are primarily reflected in the following three aspects: 1) existing studies usually assume that the spatial neighbors and time windows are globally fixed, which indicates that once the number of road segments m associated with the predicted road segment and the length of the time window n are determined, they do not change

in the spatiotemporal range. Considering the dynamic character-istics of an urban road network, traffic flow in the road network is not a static point but is a moving process from one location to another location. The spatial neighbors of the road segment primarily rely on the current traffic conditions. The number of spatial neighbors is very small if traffic congestion exists but is large during flat peak periods [5]. From the perspective of ur-ban road network heterogeneity, the number of relevant road segments for different road segments also differs; thus, sharing parameter m is difficult in the entire spatial range [29]. The selec-tion of a time window based on a time series is used to determine the length of the historical traffic data to match similar traffic patterns. The traffic data in the historical time step and the cur-rent time step must be relevant in the selection process [18]. Due to the dynamic and heterogeneous nature of the road network, even the same road segment, a significant difference is observed in the time series of traffic data in different time periods (such as morning and evening peak periods). That causes the selection of the time window to be dynamic [8]. Thus, the spatial neighbors and time windows that dynamically change over time and space are not easily described with globally fixed spatiotemporal state matrices; thus, there is a need for a dynamic spatiotemporal KNN model to adapt to the characteristics of traffic changes. 2) Existing research considers that different historical data for different time periods have different contributions to the prediction of future traffic conditions. When calculating the distance between two state spaces, the weight distance criterion is usually adopted to assign different weights to each component in the state space. The closer the time window is to the predicted time, the larger the allocated weight; the closer the spatial distance is to the predicted road segment, the greater the assigned weight [3]. However, dy-namic changes in the spatial neighbor and the time window not only affect the dimension of the space-time matrix but also cause the intensity of the correlation among different positions to dy-namically change over time. Therefore, the influence of different components of traffic data is difficult to characterize with global fixed spatiotemporal weight matrix. 3) To determine the value of the number of similar state spaces K, researchers usually employ a cross-validation method to select a suitable value, then share in the entire range of space and time[26] [28]. Due to the difference in traffic patterns in the different time periods and space loca-tions, the global fixed value of K cannot adapt to the dynamic and heterogeneous nature of a road network.

The key to short-term traffic forecasting models is the effective use of the potential spatiotemporal dependencies in the traffic data. The existing KNN models usually assume that the traffic change is a static point process and often disregard its important dynamics and heterogeneous characteristics. As a result, the structure of the prediction model is usually globally fixed in time and space, including the globally fixed spatial neighbor, time window, spatiotemporal weights, and spatiotemporal parameters, such as the traditional KNN model and the spatiotemporal KNN model.

In this paper, we propose a dynamic spatiotemporal KNN model (D-ST-KNN) for short-term traffic prediction considering spatial heterogeneity and temporal non-stationarity of city traf-fic. First, we investigated the autocorrelation of road traffic to determine the time window required for the traffic data. Second, we used the cross-correlation among different road segments to analyze the spatiotemporal dependencies of traffic and build a dynamic spatial neighbor for each road segment. The dynamic spatiotemporal state matrix is obtained by the dynamic spatial

neighbor and the dynamic time window instead of the traditional time series or the static spatiotemporal matrix to characterize the state space. Finally, we introduced the dynamic spatiotempo-ral weight, dynamic spatiotemporal parameters, and Gaussian weight function to improve the KNN model to adapt to the dy-namic and heterogeneous characteristics of the traffic.

The remainder of this paper is organized as follows: Section 2 proposes a D-ST-KNN model that considers the spatial hetero-geneity and temporal non-stationarity of city road traffic. The construction of the dynamic spatiotemporal state matrix, weights, and other parameters are also introduced in this section. In Sec-tion 3, the dynamic characteristics, prediction performance, and computational efficiency of the presented model are compre-hensively validated. The experimental results are also discussed. Section 4 concludes the paper and provides an outlook of future work.

2 METHODOLOGY

In this section, we propose a D-ST-KNN model. Our method is divided into five phases: the data bucket partition, state space definition, distance function definition, optimal neighbor selec-tion, and prediction function definition, which corresponds to Sections 2.1-2.5. First, considering the dynamic nature of traffic, the original spatiotemporal data sets are partitioned according to different time periods to form different data buckets. Second, considering the spatial heterogeneity, each segment of a data bucket is separately processed, and the appropriate spatial neigh-bors and time windows are selected. The spatiotemporal state matrix is constructed to describe the traffic conditions. Then, we introduce the spatiotemporal weight matrix to define the dis-tance function and measure the distance between the current spatiotemporal state matrix and the historical spatiotemporal state matrix to select the K nearest neighbors. Finally, we inte-grate these neighbors to obtain the predicted value of the target road segment.

2.1 Data bucket

Considering the non-stationarity and periodicity of traffic data, there are significant differences in the traffic characteristics among different time periods, such as the morning peak period, inter-peak period, and evening peak period. In the same period, the traffic data of same road segment has statistical homogeneity and the traffic pattern tends to be stable with periodic changes, such as different days for the morning peak period, which results in the spatial neighbor, the time window, and spatiotemporal parameters that can be shared. Therefore, we divide the origi-nal traffic data $\{vol_t^{L_j}, j \in [1, N], t \in [t_0, t_c]\}$ into different time periods to describe the homogeneity in same time period and dynamics in different time periods, where t_0 and t_c represent the start time step and the current time step of the time series, and L_j denotes the jth road segment.

In the study of urban traffic modeling and prediction, to distin-guish the difference among the traffic characteristics in different time periods, [24] divided a day into six time periods (period 1: midnight-6:30 am; period 2: 6:30-10:00; period 3: 10:00-13:30; pe-riod 4:13:30-17:00; period 5:17:00-20:30; period 6:20:30-midnight). The test reveals that the partition is statistically acceptable. Based on this analysis and according to the same strategy, the original traffic data are divided into M different time periods (M= 6) ac-cording to the time dimension, which corresponds to different data buckets. Assuming that the entire traffic data set is BK, the

data bucket division must be satisfied:

$$
\begin{cases}
BK = bk_1 \cup bk_2 \cup ... \cup bk_M \\
bk_i = \{vol_t^{L_j} | 1 \le j \le N, \forall t \in [t_a^{bk_i}, t_b^{bk_i})\} \\
bk_i \cap bk_o = \phi
\end{cases} \quad (1)
$$

where $i \in [1, M], o \in [1, M], i \ne o$, bk_i is the ith bucket (i.e., bucket 1), and $vol_t^{L_j}$ is the traffic data of road segment L_j at time step t. $t \in [t_a^{bk_i}, t_b^{bk_i})$ indicates that time step t is within the corresponding time period of the ith bucket (i.e.,[0:00-6:30], [6:30-10:00)). L_j denotes the jth road segment (i.e., Link 1), and N is the total number of road segments. Note that dividing the original traffic data into different buckets at the pre-processing stage does not have any impact on the analyses and conclusions in this study because the same partitioning strategy were used for all the algorithms that are evaluated.

2.2 Dynamic spatiotemporal state matrix

2.2.1 Dynamic spatial neighborhoods. The dynamic spatial neighborhood is used to determine how the traffic conditions of the predicted road segment are affected by the surrounding road segments in different buckets to determine the correlation among road segments. The traditional method usually calculates the correlation coefficients between the time series of the predicted road segments and the time series of other road segments and sets the threshold to select the relevant road segments [3]. Considering that a road network has multiple internal and external factors, such as the influence of traffic lights, the impact of surrounding road segments on predicted road segments has a certain degree of lag. Therefore, the delayed spatiotemporal relationships cannot be exactly expressed by correlation coefficients. The cross-correlation function is a delayed version of the correlation coefficient function, which measures the correlation coefficients of two time series at a specific lag [14]; therefore, it is more suitable for describing the spatiotemporal dependence of traffic.

Assume that bk_i is the bucket of the predicted road segment L_j at time step t, and $t \in [t_a^{bk_i}, t_b^{bk_i})$. Given the surrounding road segments L_v, the time series of the traffic data for two road segments can be expressed as $U = \{vol_t^{L_j} | \forall t \in [t_a^{bk_i}, t_b^{bk_i})\}$, $Z = \{vol_t^{L_v} | \forall t \in [t_a^{bk_i}, t_b^{bk_i})\}, j \in [1, N], v \in [1, N]$, and their cross-correlation at lag φ is defined as follows:

$$
\begin{cases}
ccf_{u,z}^{bk_i}(\varphi) = \frac{\gamma_{u,z}^{bk_i}(\varphi)}{\alpha_u \sigma_z}, \varphi = 0, \pm 1, \pm 2, \cdots, \\
\gamma_{u,z}^{bk_i}(\varphi) = E\left[(u_t - \mu_u)(z_{t+\varphi} - u_z)\right] \\
\alpha_u = \sqrt{\sum (u_t - \mu_u)^2} \\
\sigma_z = \sqrt{\sum (z_{t+\varphi} - u_z)^2}
\end{cases} \quad (2)
$$

where $\gamma_{u,z}^{bk_i}(\varphi)$ is the correlation coefficient between time series U and time series Z at lag φ in bucket bk_i, μ_u and u_z are the mean values of U and Z, respectively, and σ_u and σ_z are the standard deviations of U and Z, respectively.

In this definition, the cross-correlation function can be regarded as a function of lag, and the lag value that makes the cross-correlation function obtain the maximum value is the average delay time of the surrounding segments to the predicted road segment [29]. The formal definition is expressed as

$$
\psi_{bk_i}^{L_v} = \left| \underset{\varphi}{argmax} \left(ccf_{u,z}^{bk_i}(\varphi) \right) \right|, v \in [1, N] \quad (3)
$$

where $\psi_{bk_i}^{L_v}$ is the lag value that maximizes cross-correlation of the surrounding road segment L_v to the predicted road segment in bk_i, and $\psi_{bk_i}^{L_v}$ describes the maximum impact time range of the surrounding segments in different buckets on the predicted road segment, which can be employed for efficient selection of spatial neighbors. Consider the predicted road segment L_j in bk_i and its predicted time interval Δt. When the surrounding road segments deliver the traffic flow to the predicted road segments within a given time interval, they influence the predicted road segments, and the road segments beyond this time interval are excluded. Its formal definition is expressed as

$$
R_{bk_i}^{L_j} \leftarrow \left\{ L_v | \forall 0 \le \psi_{bk_i}^{L_v} \le \Delta t, v \in [1, N] \right\} \quad (4)
$$

where $R_{bk_i}^{L_j}$ is the set of spatial neighbors of the jth road segment in the ith bucket.

2.2.2 Dynamic time windows. Considering that the selection of the time window is based on the time series of the predicted road segment, we can select n historical traffic data that have a correlation with the predicted road segment. The autocorrelation function is usually employed to measure the correlation between the time series and its delayed version; thus, it can be used for the selection of the time window, i.e., the lag in which the prediction error is minimized can be set as the window size. Note that the lag in the autocorrelation function describes the delay effect of the time series, and the lag described in Section 2.2.1 is used to characterize the delay effect between different time series. Given the time series of the jth road segment L_j in bk_i, $U = \{vol_t^{L_j} | \forall t \in [t_a^{bk_i}, t_b^{bk_i})\}$, the autocorrelation function $\rho_{bk_i}^{L_j}(\delta)$ can be defined as follows:

$$
\rho_{bk_i}^{L_j}(\delta) = \frac{E[(u_t - \mu_u)(u_{t-\delta} - \mu_u)]}{\sigma_u^2}, \delta = 0, 1, 2, \cdots, \quad (5)
$$

Using the autocorrelation function to set the time window entails three steps. First, consider the computational limitations, it is necessary to determine the maximum range of lag. Second, within the range, the parameters of the predictive model are fixed, and cross-validation is performed with different lags. This strategy is based on the fact that the value of the traffic data has a significant correlation within the maximum lag range. Finally, the lag that minimizes the prediction error is chosen as the optimal time window.

2.3 Dynamic spatiotemporal weights

Considering the traffic conditions have significant differences at different time intervals, which results in a change in the spatiotemporal weights with time; the historical data of different time and space will influence the future traffic conditions by a different degree. The dimension of the spatiotemporal weight is related to the spatiotemporal state matrix, and the dynamic change in the spatiotemporal matrix causes the dimension of the spatiotemporal weight matrix to change with different time periods. Based on the traditional weight distance function, we introduce a dynamic spatiotemporal weight in the distance function and optimize the weight distance function to adapt the nearest neighbor similarity measure of the dynamic spatiotemporal matrix.

In the temporal dimension, we use the time interval length (i.e., 5 min interval) to characterize the contribution of different time steps. In the spatial dimension, the spatial correlation (such as cross-correlation) is used to characterize the influence of different

spatial distances. The construction method is described as follows: assume that the predicted road segment L_j at the current time step t_c is in data bucket bk_i and the dimension of the spatiotemporal state matrix is $m^{L_j}_{bk_i} \times n^{L_j}_{bk_i}$, which is determined by the method provided in Section 2.2. Then, the spatiotemporal state matrix of the current time step can be expressed as $\chi^{L_j}_{t_c}\left(m^{L_j}_{bk_i}, n^{L_j}_{bk_i}\right)$. The spatiotemporal matrix of the historical time step h_i can be defined as $\chi^{L_j}_{h_i}\left(m^{L_j}_{bk_i}, n^{L_j}_{bk_i}\right)$, where $m^{L_j}_{bk_i}$ is the spatial dimension of the spatiotemporal state matrix of the jth predicted road segment in the ith bucket, which is related to the number of elements in the set of spatial neighbors $R^{L_j}_{bk_i}$. Moreover, $n^{L_j}_{bk_i}$ is the temporal dimension of the spatiotemporal state matrix of the jth predicted road segment in the ith bucket, which is the size of the time window. The time-weighted matrix is defined as $W^{bk_i}_t$, and the space-weighted matrix is defined as $W^{bk_i}_s$. The corresponding elements are $w^{bk_i}_t(ti, tj), ti \in [1, n^{L_j}_{bk_i}], tj \in [1, n^{L_j}_{bk_i}]$ and $w^{bk_i}_s(si, sj), si \in [1, m^{L_j}_{bk_i}], sj \in [1, m^{L_j}_{bk_i}]$, which represent the time weight value and space weight value, respectively, assigned to the jth predicted road segment in the ith bucket. The weight distribution is as follows:

$$w^{bk_i}_t(ti, tj) = \begin{cases} ti \left/ \sum_{ti=1}^{n^{L_j}_{bk_i}} ti \right., ti = tj \\ 0, ti \neq tj \end{cases} \quad (6)$$

$$w^{bk_i}_s(si, sj) = \begin{cases} ccf^{si}_{L_v, L_j} \left/ \sum_{si=1}^{m^{L_j}_{bk_i}} ccf^{si}_{L_v, L_j} \right., si = sj \\ 0, si \neq sj \end{cases} \quad (7)$$

In this definition, the temporal and spatial weights are linearly distributed according to the proximity of the current time step and the predicted road segments. $ccf^{si}_{L_v, L_j}$ is the cross-correlation between the time series of the si spatial neighbor (whose road segment is L_v) and the predicted road segment L_j. The closer the value is to the predicted time, the greater the weight of the allocation; the greater the relation to the space of the predicted road segment, the greater the weight. By introducing spatiotemporal weights into the original spatiotemporal matrix, the spatiotemporal-weighted state matrices of the current time step $\Gamma^{L_j}_{t_c}$ and the spatiotemporal-weighted state matrices of the historical time step $\Gamma^{L_j}_{h_i}$ are denoted by the following:

$$\Gamma^{L_j}_{t_c} = W^{bk_i}_s \times \chi^{L_j}_{t_c}\left(m^{L_j}_{bk_i}, n^{L_j}_{bk_i}\right) \times W^{bk_i}_t \quad (8)$$

$$\Gamma^{L_j}_{h_i} = W^{bk_i}_s \times \chi^{L_j}_{h_i}\left(m^{L_j}_{bk_i}, n^{L_j}_{bk_i}\right) \times W^{bk_i}_t \quad (9)$$

By calculating the distance $d_{bk_i}(\Gamma^{L_j}_{t_c}, \Gamma^{L_j}_{h_i})$ between the historical spatiotemporal state matrix and the current spatiotemporal state matrix, candidate neighbors can be selected. The formula is expressed as

$$d_{bk_i}\left(\Gamma^{L_j}_{t_c}, \Gamma^{L_j}_{h_i}\right) = \sqrt{trac\left(\left(\Gamma^{L_j}_{t_c} - \Gamma^{L_j}_{h_i}\right) \times \left(\Gamma^{L_j}_{t_c} - \Gamma^{L_j}_{h_i}\right)'\right)} \quad (10)$$

where $trac$ represents the trace of the matrix.

2.4 Dynamic spatiotemporal parameters

In the KNN model, the spatiotemporal parameters include the K values and the parameters introduced during the method construction (such as the prediction generation functions). The reasonableness of the parameters has substantial influence on the

prediction accuracy of the model. The K value is primarily employed to determine the number of candidate neighbors. If the K value is too small, the model becomes more complex and overfitting is possible. If the K value is too large, the model is simpler and under-fitting is possible. Considering that the selection of the K value is significantly influenced by the finite sample nature of the problem, the assignment of its values is usually performed by cross-validation to select the K value that minimizes the model error [27].

The existing methods usually assume that the K value is globally fixed. When the K value is determined, it is shared throughout the entire space and time. In contrast to the existing method, the selection of the K value in the D-ST-KNN model considers the characteristics of dynamic changes of traffic. Instead of setting a global fixed K value, we can select the optimal K value for different buckets, i.e., $K_{bk_i}, bk_i \in BK, i \in [1, M]$.

To verify these assumptions, we use cross-validation to set the range of K to $[1, 40]$ and test the effect of different K values on MAPE of the model in different buckets, as shown in Fig. 1.
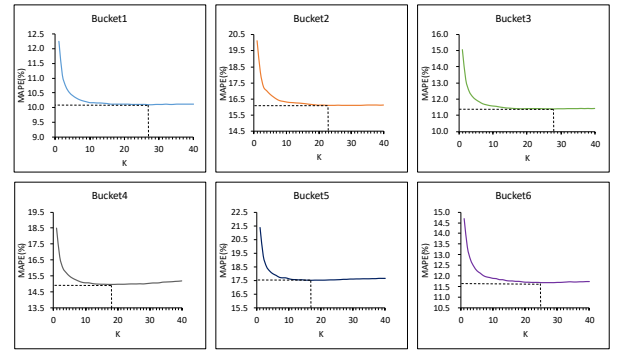

Figure 1: Impact of the number of candidate neighbors K_{bk_i} on the MAPEs of different data buckets.

As the K value increases, the prediction error is gradually reduced. When the K value attains a certain value, the error of the model begins to stabilize. Thus, the optimal K value for each bucket can be determined (i.e., $K_{bk_1} = 27, K_{bk_2} = 23$). Compared with different buckets, the K values dynamically vary with different time periods. The global fixed K value has difficulty describing the dynamic change in traffic. Therefore, the dynamic K value proposed in this paper is reasonable. The parameters of the D-ST-KNN model also contain the parameters introduced by the predicted generation function (refer to Section 2.5). The calibration method of the parameter is shown in Section 3.2.

2.5 Predictive function

Due to the spatiotemporal state space, the spatiotemporal weight, and the spatiotemporal parameters dynamically change with different buckets; to adapt to this change, the predictive generation function should also dynamically change. This paper transforms the four types of traditional weight distribution methods to enable them to adapt to the dynamics of traffic, including the inverse distance weight [23], rank-based weight[11][13], and Gaussian weight [3]. Selecting the best prediction function by comparing the performance of different predictive functions (refer to Section 3.2). Note that the weight referred to in this section is expressed as the weight assigned by the candidate neighbor, whereas the weight in Section 2.3 represents the weight matrix of the weights assigned to each element in the spatiotemporal state matrix.

Assuming that $d_{bk_i}^k$ is the distance between the kth candidate neighbor and the predicted road segment in the ith bucket obtained by formula (10), $\widehat{vol_{t_c+1}^{L_j}}$ the predicted value of the predicted road segment L_j at time step $t_c + 1$ is defined as

$$\widehat{vol_{t_c+1}^{L_j}} = \frac{\sum_{k=1}^{K_{bk_i}} vol_{h_i+1}^{L_j}(k) \times \varphi_{bk_i}^{L_j}(k)}{\sum_{k=1}^{K_{bk_i}} \varphi_{bk_i}^{L_j}(k)} \qquad (11)$$

where $t_c \in [t_a^{bk_i}, t_b^{bk_i}]$ is used to map the current time step into the corresponding bucket, is used to determine the number of candidate neighbors for the corresponding bucket, $vol_{h_i+1}^{L_j}(k)$ represents the traffic data of the kth candidate neighbor, and $h_i \in [t_a^{bk_i}, t_b^{bk_i}]$; and $\varphi_{bk_i}^{L_j}(k)$ and represent the weight of the kth neighbor of the jth predicted road segment in the ith bucket. The form is defined as follows:

$$\varphi_{bk_i}^{L_j}(k) = \begin{cases} \frac{1}{K_{bk_i}} \\ \frac{1}{d_{bk_i}^k} \\ (K_{bk_i} - r_q + 1)^2 \\ \frac{1}{4\pi a_{bk_i}} exp(-\frac{\left|d_{bk_i}^k\right|^2}{4a_{bk_i}^2}) \end{cases} \qquad (12)$$

Formula (12) corresponds to equal weights, inverse distance weights, the rank-based weight and the Gaussian weight, where r_q represents the order of the qth candidate neighbors, and a_{bk_i} is the spatiotemporal parameter whose value is similar to the value of the previously discussed spatiotemporal parameter K, which dynamically values with different time periods. The corresponding parameter calibration is shown in Section 3.2.

2.6 Accuracy metrics

Three criteria are selected to verify the prediction accuracy of the D-ST-KNN model, namely, mean absolute error (MAE), mean absolute percentage error (MAPE) and root-mean-square error (RMSE). These indicators depict the essential characteristics of errors from different perspectives. The RMSE indicates a fluctuation in the error of the prediction model, and the MAPE indicates the difference between the predicted and the actual traffic data. In contrast, the MAE and RMSE provide a measure of the similarity between the predicted and the actual traffic data [12]. The MAE, MAPE, and RMSE are defined as follows:

$$MAE = \frac{1}{M \times N \times S} \sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{s=1}^{S} \left| vol_{t_c+1}^{L_j}(s) - \widehat{vol_{t_c+1}^{L_j}}(s) \right| \qquad (13)$$

$$MAPE = \sqrt{\frac{1}{M \times N \times S} \sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{s=1}^{S} \frac{\left(vol_{t_c+1}^{L_j}(s) - \widehat{vol_{t_c+1}^{L_j}}(s) \right)}{vol_{t_c+1}^{L_j}(s)}} \qquad (14)$$

$$RMSE = \sqrt{\frac{1}{M \times N \times S} \sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{s=1}^{S} \left(vol_{t_c+1}^{L_j}(s) - \widehat{vol_{t_c+1}^{L_j}}(s) \right)^2} \qquad (15)$$

where M is the number of buckets $M = 6$, N is the number of predicted road segments, S is the number of test samples, $vol_{t_c+1}^{L_j}(s)$ and $\widehat{vol_{t_c+1}^{L_j}}(s)$ indicate the actual traffic data and the predicted traffic data at the next time step of the jth predicted road segment at the current time step, and s indicates the sth test sample in the ith bucket.

3 EXPERIMENTS

3.1 Data preparation

In this study, two different data sets are used to evaluate the performance of the prediction model. The first data set is PeMS, which is a high-quality data set with open access. PeMS is extensively applied in the field of traffic prediction. The traffic speed data from 59 consecutive locations on the US 101 freeway from PeMS were downloaded for a total of 60 days; the time period is August 15, 2016, to October 14, 2016 and time interval is 5 min (as shown in Table 1). Each detector represents a position; the positional distribution is shown in Fig. 2. The second data set is the floating car trajectory data obtained from the Beijing road network, which is generated from more than 50,000 vehicles equipped with GPS. The frequency of data acquisition is 5 min, and the time period is March 1, 2012, to April 30, 2012 (as shown in Table 1). In this study, a representative region that contains 30 road segments is used for the experiment with the position distribution shown in Fig. 2. In the two data sets, the last ten days are used as the test data to evaluate the accuracy of the model. The remaining days of data are employed as training data to construct the historical database of the predicting model.

In addition, we normalize the original traffic data and use the ratio of the average traffic speed to the maximum speed limit of each road segment to express the traffic conditions of the road segment. The formal expression is as follows:

$$\widehat{v_{i,t}} = \frac{v_{i,t}}{f_{i,\max}}, \quad i \in [1, N], \ t \in [t_0, t_c] \qquad (16)$$

where $\widehat{v_{l,t}}$ is the normalized speed of the ith road segment at time step t, $v_{i,t}$ is the real average speed data of the road segment, and $f_{i,max}$ is the speed limit for the ith road segment.

Table 1: Description of the experimental data sets

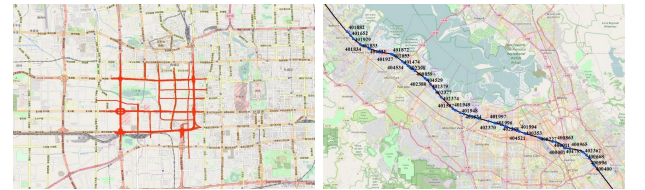| Data set	PeMS	Beijing
Time span	8/15/2016-10/14/2016	3/1/2012-4/30/2012
Time interval	5 min	5 min
Number of links	59	30

Figure 2: Spatial distribution of traffic data in the Beijing and PeMS data set.

3.2 Variable estimation

3.2.1 Determining the optimal distance function. The distance function is used to measure the similarity among the spatiotemporal state matrices to obtain the historical spatiotemporal matrix, which is similar to the spatiotemporal state matrix of the target road segment. Fig. 3 shows the performance differences of the distance function constructed with different weights. The traditional method directly calculates the Euclidean distance between two spatiotemporal state matrices, which treats the elements in the space state matrix equally. The influence of the historical traffic conditions of different time and space distribution on the prediction of future traffic conditions is difficult to describe, which

yields the lowest performance. The distance function constructed by the Gaussian function assigns weights in the time dimension and space dimension; thus, the performance of the prediction model is significantly improved. However, this method requires additional introduction of the time-weighted parameter α_1 and the space-weighted parameter α_2 in the construction process, which makes calibration of its parameters and the global optimal combination of parameters difficult. We adopt a similar strategy that uses the linear time distribution weight in the time dimension and the spatial correlation between the surrounding road segments and the target road segment to assign weights in spatial dimensions. Then, a dynamic spatiotemporal weight assignment method is constructed that does not require any additional parameters. The dynamic weight distribution has the lowest MAPE, RMSE and MAE, which reflects the high efficiency of the method compared to that of the other two weight distribution methods.
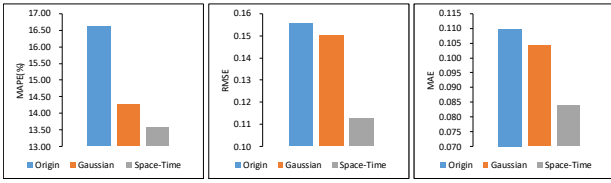

Figure 3: Comparison of different distance functions

3.2.2 Determining the optimal predictive function. Based on the discussion in the previous sections, we transform four types of weight distribution methods, including equal weight, inverse distance weight, rank-based weight, and Gaussian weight, which are used to integrate the candidate neighbors to obtain the final predicted value. In the process of cross-validation, we fix the other parameters of the model, such as K_{bk_i} and a_{bk_i}, and calculate the influence of different weight distribution methods on the prediction accuracy of the D-ST-KNN model to obtain the average error of the entire test data set for different weight distribution methods. The results are shown in Fig. 4. The MAPE, RMSE and MAE of the Gaussian weight method are lower than the MAPE, RMSE and MAE of the other three weight distribution methods. In the D-ST-KNN model, we employ the Gaussian function as the weight distribution method for candidate neighbors.
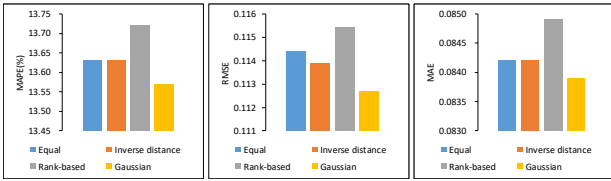

Figure 4: Comparison of different weight allocation methods.

3.2.3 Calibrating hyper-parameters In the D-ST-KNN model, the hyper-parameters primarily include the number of candidate neighbors K_{bk_i} and the Gaussian weight parameter a_{bk_i}. In the parameter calibration process, to find the best combination of K_{bk_i} and a_{bk_i} that enables the prediction model to obtain the minimum MAPE, we set the range of K_{bk_i} to $[1, 40]$ and the range of a_{bk_i} to $[0.001, 0.04]$. We apply the cross-validation method to obtain the optimal combination of the parameters for each bucket. The effect of parameter variation on the prediction accuracy of the D-ST-KNN model can be tested by fixing other parameters of the model. For example, we can fix the values of a_{bk_i} and test the

performance of the prediction model changes with K_{bk_i} (refer to Section 2.4). Because the impact of parameter K_{bk_i} on the prediction performance was discussed in Section 2.4, this section focuses on the calibration of parameter a_{bk_i}.

Fig. 5 shows the impact of changes in a_{bk_i} on the performance of the D-ST-KNN model in different buckets. The trend in Fig. 5 reveals that the value of a_{bk_i} has a significant influence on the prediction performance. For the minimum a_{bk_i}, the prediction error of the model attains the maximum a_{bk_i}. As a_{bk_i} increases, the prediction error gradually decreases and begins to stabilize. We compare the variation of the parameters among the different buckets. For example, in bucket 1, the optimal value of a_{bk_1} is 0.017, whereas the optimal value of a_{bk_2} in bucket 2 is 0.015. The value of a_{bk_i} also changes dynamically over time. Considering that K_{bk_i} also changes dynamically with time, the parameters of the D-ST-KNN model change with time. The calibration results of the entire model are listed in Table 2, and the values of K_{bk_i} are shown in Fig. 5. In this analysis, setting the global fixed parameters is unreasonable when constructing the prediction model. We propose the concept of the data bucket, and the prediction model is constructed in different time periods, which causes the model parameters to change with the time period to adapt to the dynamic nature of traffic.

Table 2: Calibration results of the model parameters

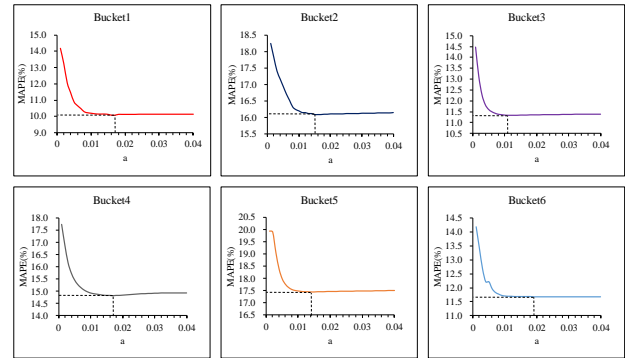| Bucket	Parameters	
	K_{bk_i}	a_{bk_i}
Bucket 1	27	0.017
Bucket 2	23	0.015
Bucket 3	28	0.011
Bucket 4	18	0.017
Bucket 5	17	0.014
Bucket 6	25	0.019

Figure 5: Impact of the weight parameter a_{bk_i} on MAPEs for different data buckets.

3.3 Accuracy evaluation

3.3.1 Overall results. Based on the variable estimation, we compare our model with several existing traffic prediction models, including the historical average model (HA), Elman neural network (Elman-NN) [9], traditional KNN model (Original-KNN), and spatiotemporal KNN model (ST-KNN). Fig. 6 shows the prediction performance of different models. The HA model, the Elman-NN model, and the Original-KNN model regard the problem of the traffic prediction as a simple time series problem and disregard

the influence of the spatial factors on the predicted road segment. Therefore, their prediction performance is lower than the prediction performance of the ST-KNN model and the D-ST-KNN model proposed in this paper by comparing the values of MAPE. The ST-KNN model introduces the spatiotemporal state matrix, which improves the prediction performance of the model. However, this matrix ignores the spatial heterogeneity and the temporal non-stationarity of the road network and cannot describe the essential characteristics of the traffic dynamics using a static ST-KNN model (including global fixed spatiotemporal matrix and global fixed parameters). The D-ST-KNN model constructs models for different time periods by introducing the concept of data buckets. Simultaneously, the dynamic space neighbor, dynamic time window, dynamic spatiotemporal weight, and dynamic spatiotemporal parameters are introduced to construct the D-ST-KNN model, which can adequately adapt to the dynamic changes of traffic conditions. The experimental results indicate that the D-ST-KNN model proposed in this paper is superior to other models.
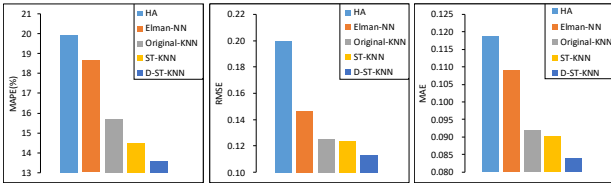

Figure 6: Accuracy comparison of different models in the Beijing data set.

3.3.2 Local results. To further evaluate the performance of the D-ST-KNN model, we compare the MAPEs of different models in different data buckets by averaging the prediction performance of different road segments in a single bucket. The experimental results are displayed in Fig. 7. In terms of overall trends, the performance of different models corresponds to the degree of congestion of the traffic conditions. For example, in bucket 1, bucket 3, and bucket 6, all models have a lower MAPE than other buckets because the time periods that correspond to the three data buckets are midnight-6:30 am, 10:00-13:30, and 20:30-midnight. The traffic in Beijing during these three time periods belongs to the flat peak period, and road traffic has low congestion and exhibits regular changes. In buckets 2, 4, and 5, all models achieve a relatively poor performance. Bucket 2 corresponds to the time period of 6:30-10:00, bucket 4 corresponds to the time period of 13:30-17:00 and bucket 5 corresponds to the time period of 17:00-20:30. These time periods correspond to the peak period in Beijing. The changes of traffic conditions during these time periods are more complicated than the traffic conditions of the other buckets. In addition, in terms of the performance of different models in a single data bucket, the prediction trend of different models was similar to those of the overall results. For example, in bucket 1, the ST-KNN and D-ST-KNN models perform better than the HA, Elman-NN, and Original-KNN models, which is due to the benefits of the introduction of spatial factors. However, the D-ST-KNN model considers the spatial heterogeneity and temporal non-stationarity of road networks to adapt to the dynamic characteristics of traffic, making the model performance better than other models in all time periods, especially in the peak period. This also explains why the D-ST-KNN model is superior to the other models in the overall result.
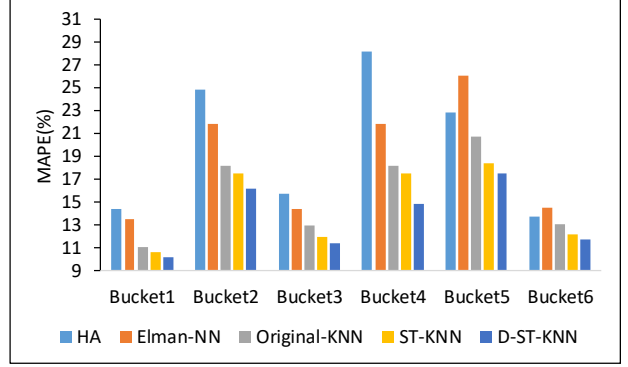

Figure 7: Accuracy comparison of the D-ST-KNN model on MAPEs for different data buckets.

3.4 Generalization ability evaluation

To evaluate the generalization ability of the D-ST-KNN model, we fix all parameters of the model and compare the performance of the different methods with the test data set from PeMS; the experimental results are shown in Fig. 8. The results indicate that the prediction accuracy of the D-ST-KNN model on the PeMS data set is significantly improved compared with that of the Beijing floating car data set. The data quality of the PeMS data set is relatively complete, and the data collection area is the expressway. Compared with the traffic conditions of the urban road network, the traffic mode is relatively simple with minimal changes, which enables the prediction model to easily represent the regular traffic pattern characteristics. However, the D-ST-KNN model maintains the same prediction trend; in all predicted models, its MAPE, RMSE, and MAE are lower than the other models, which exhibit excellent predictive performance and generalization ability.
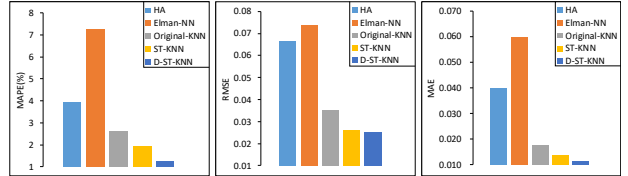

Figure 8: Performance comparison of different models in the PeMS data set.

4 SUMMARY AND FUTURE WORK

In this paper, we propose a D-ST-KNN model for short-term traffic prediction. The proposed model considers the spatial heterogeneity and temporal non-stationarity of road networks to adapt to the dynamic characteristics of traffic, including dynamic spatial neighbors, time windows, spatiotemporal weights, and spatiotemporal parameters. With cross-correlation and autocorrelation function computation, the automatic selections of spatial neighbors and the time window are realized, which efficiently solve the dimensionality disaster problem encountered in the existing KNN models. The spatiotemporal weights are integrated into a distance function to help identify candidate neighbors. Time variable parameters are also introduced, including the dynamic number of candidate neighbors and dynamic weight allocation parameters, to further adapt to the dynamic and heterogeneous nature of road networks.

Using real traffic data collected from city roads and inter-city expressways, we calculate the number of spatial neighbors and the time window size of each road segments, which reflects the distinct heterogeneity and non-stationarity of urban road traffic. Then, we validate the performance of the proposed D-ST-KNN model with comparisons to HA, Elman-NN, traditional KNN and spatiotemporal KNN models. The experimental results indicate that the D-ST-KNN model has a higher accuracy on short-term traffic prediction than the existing models. In addition, we explore the local performance of different models in different data buckets and find that all models correspond to the degree of traffic congestion, and the D-ST-KNN model performs better than other models in all time periods, especially in the morning period and evening peak period. To summarize, compared with the existing models, the proposed D-ST-KNN model significantly improves the accuracy of short-term traffic prediction. Furthermore, we compare the performance of different models using the actual traffic data collected from PeMS. The D-ST-KNN model also achieves the best performance, which verifies the generalization ability of the proposed model.

In the follow-up study, the following problems need to be investigated to further improve the D-ST-KNN model. The D-ST-KNN model behaves slightly differently in peak and off-peak time periods. Further improvement of the model performance during peak hours will be a constant challenge. Moreover, a multi-threaded approach could be used to improve the efficiency of D-ST-KNN. A parallel P-D-ST-KNN model on an existing parallel computing framework is expected to alleviate the pressure of real-time computation.

5 ACKNOWLEDGMENTS

REFERENCES

[1] J Scott Armstrong. 2006. Findings from evidence-based forecasting: Methods for reducing forecast error. *International Journal of Forecasting* 22, 3 (2006), 583–598.

[2] Brenda I. Bustillos and Yi Chang Chiu. 2011. Real-Time Freeway-Experienced Travel Time Prediction Using N -Curve and k Nearest Neighbor Methods. *Transportation Research Record Journal of the Transportation Research Board* 2243, -1 (2011), 127–137.

[3] Pinlong Cai, Yunpeng Wang, Guangquan Lu, Peng Chen, Chuan Ding, and Jianping Sun. 2016. A spatiotemporal correlative k -nearest neighbor model for short-term traffic multistep forecasting. *Transportation Research Part C Emerging Technologies* 62 (2016), 21–34.

[4] H. Chang, Y. Lee, B. Yoon, and S. Baek. 2012. Dynamic near-term traffic flow prediction: systemoriented approach based on past experiences. *Iet Intelligent Transport Systems* 6, 3 (2012), 292–305.

[5] Tao Cheng, James Haworth, and Jiaqiu Wang. 2012. Spatio-temporal autocorrelation of road network data. *Journal of Geographical Systems* 14, 4 (2012), 389–413.

[6] Tao Cheng, Jiaqiu Wang, James Haworth, Benjamin Heydecker, and Andy Chow. 2014. A Dynamic Spatial Weight Matrix and Localized SpaceĺCTime Autoregressive Integrated Moving Average for Network Modeling. *Geographical Analysis* 46, 1 (2014), 75ĺC97.

[7] Stephen Clark. 2003. Traffic Prediction Using Multivariate Nonparametric Regression. *Journal of Transportation Engineering* 129, 2 (2003), 161–168.

[8] Peibo Duan, Guoqiang Mao, Shangbo Wang, Changsheng Zhang, and Bin Zhang. 2016. STARIMA-based Traffic Prediction with Time-varying Lags. (2016).

[9] Jeffrey L Elman. 1990. Finding structure in time. *Cognitive Science* 14, 2 (1990), 179–211.

[10] Gaetano Fusco, Chiara Colombaroni, and Natalia Isaenko. 2016. Short-term speed predictions exploiting big data on large urban road networks. *Transportation Research Part C: Emerging Technologies* 73 (2016), 183–201.

[11] Kamran Ghavamifar. 2009. A decision support system for project delivery method selection in the transit industry. (2009).

[12] Filmon G. Habtemichael and Mecit Cetin. 2016. Short-term traffic flow rate forecasting based on identifying similar traffic patterns. *Transportation Research Part C* 66 (2016), 61–78.

[13] Filmon G. Habtemichael, Mecit Cetin, and Khairul A. Anuar. 2015. METHODOLOGY FOR QUANTIFYING INCIDENT-INDUCED DELAYS ON FREEWAYS BY GROUPING SIMILAR TRAFFIC PATTERNS. *Transportation Research Record Journal of the Transportation Research Board* (2015).

[14] James Douglas Hamilton. 1994. *Time series analysis*. 401–409 pages.

[15] Haikun Hong, Wenhao Huang, Xiabing Zhou, Sizhen Du, Kaigui Bian, and Kunqing Xie. 2016. Short-term traffic flow forecasting: Multi-metric KNN with related station discovery. In *International Conference on Fuzzy Systems and Knowledge Discovery*. 1670–1675.

[16] Xiaoyu Hou, Yisheng Wang, and Siyu Hu. 2013. Short-term Traffic Flow Forecasting based on Two-tier K-nearest Neighbor Algorithm ąî. *Procedia - Social and Behavioral Sciences* 96 (2013), 2529–2536.

[17] Myung Jiwon, Dong Kyu Kim, Seung Young Kho, and Chang Ho Park. 2011. Travel Time Prediction Using k Nearest Neighbor Method with Combined Data from Vehicle Detector System and Automatic Toll Collection System. *Transportation Research Record Journal of the Transportation Research Board* 20, 2256 (2011), 51–59.

[18] Arief Koesdwiady, Ridha Soua, and Fakhreddine Karray. 2016. Improving Traffic Flow Prediction With Weather Information in Connected Cars: A Deep Learning Approach. *IEEE Transactions on Vehicular Technology* 65, 12 (2016), 9508–9517.

[19] Shuangshuang Li, Zhen Shen, and Gang Xiong. 2012. A k-nearest neighbor locally weighted regression method for short-term traffic flow forecasting. In *International IEEE Conference on Intelligent Transportation Systems*. 1596–1601.

[20] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang. 2017. Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction. *Sensors* 17, 4 (2017).

[21] Wanli Min and Laura Wynter. 2011. Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies* 19, 4 (2011), 606–616.

[22] Brian Lee Smith. 1997. Forecasting freeway traffic flow for intelligent transportation systems application. *Transportation Research Part A* 1, 31 (1997), 61.

[23] Brian L Smith, Billy M Williams, and R Keith Oswald. 2002. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies* 10, 4 (2002), 303–321.

[24] Anthony Stathopoulos and Matthew G. Karlaftis. 2003. A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C Emerging Technologies* 11, 2 (2003), 121–135.

[25] Eleni I. Vlahogianni, Matthew G. Karlaftis, and John C. Golias. 2007. Spatio-Temporal Short-Term Urban Traffic Volume Forecasting Using Genetically Optimized Modular Networks. *Computer-Aided Civil and Infrastructure Engineering* 22, 5 (2007), 317ĺC325.

[26] Shanhua Wu, Zhongzhen Yang, Xiaocong Zhu, and Bin Yu. 2014. Improved k-nn for Short-Term Traffic Forecasting Using Temporal and Spatial Information. *Journal of Transportation Engineering* 140, 7 (2014), 04014026.

[27] Dawen Xia, Binfeng Wang, Huaqing Li, Yantao Li, and Zili Zhang. 2016. A distributed spatialĺCtemporal weighted model on MapReduce for short-term traffic flow forecasting. *Neurocomputing* 179, C (2016), 246–263.

[28] Bin Yu, Xiaolin Song, Feng Guan, Zhiming Yang, and Baozhen Yao. 2016. k-Nearest neighbor model for multiple-time-step prediction of short-term traffic condition. *Journal of Transportation Engineering* 142, 6 (2016), 04016018.

[29] Yang Yue and Anthony Gar-On Yeh. 2008. Spatiotemporal traffic-flow dependency and short-term traffic forecasting. *Environment and Planning B: Planning and Design* 35, 5 (2008), 762–771.

[30] Lun Zhang, Qian Rao, Wenchen Yang, and Meng Zhang. 2013. An Improved k-NN Nonparametric Regression-Based Short-Term Traffic Flow Forecasting Model for Urban Expressways. In *International Conference on Transportation Engineering*. 1214–1223.

[31] Zuduo Zheng and Dongcai Su. 2014. Short-term traffic volume forecasting: A k -nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm. *Transportation Research Part C Emerging Technologies* 43 (2014), 143–157.