Prediction of Spatiotemporal Evolution of Urban Traffic Emissions Based on Taxi Trajectories

Zhen-Yi Zhao[1]   Yang Cao[1]   Yu Kang[2,3]   Zhen-Yi Xu[4,5]

[1]Department of Automation, University of Science and Technology of China, Hefei 230026, China

[2]Department of Automation and State Key Laboratory of Fire Science, University of Science and Technology of China, Hefei 230026, China

[3]Institute of Advanced Technology, University of Science and Technology of China, Hefei 230088, China

[4]Department of Automation, University of Science and Technology of China, Hefei 230026, China

[5]Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230088, China

Abstract: With the rapid increase of the amount of vehicles in urban areas, the pollution of vehicle emissions is becoming more and more serious. Precise prediction of the spatiotemporal evolution of urban traffic emissions plays a great role in urban planning and policy making. Most existing methods usually focus on estimating vehicle emissions at historical or current moments which cannot well meet the demands of future planning. Recent work has started to pay attention to the evolution of vehicle emissions at future moments using multiple attributes related to emissions, however, they are not effective and efficient enough in the combination and utilization of different inputs. To address this issue, we propose a joint framework to predict the future evolution of vehicle emissions based on the GPS trajectories of taxis with a multi-channel spatiotemporal network and the motor vehicle emission simulator (MOVES) model. Specifically, we first estimate the spatial distribution matrices with GPS trajectories through map-matching algorithms. These matrices can reflect the attributes related to the traffic status of road networks such as volume, speed and acceleration. Then, our multi-channel spatiotemporal network is used to efficiently combine three key attributes (volume, speed and acceleration) through the feature sharing mechanism and generate a precise prediction of them in the future period. Finally, we adopt an MOVES model to estimate vehicle emissions by integrating several traffic factors including the predicted traffic states, road networks and the statistical information of urban vehicles. We evaluate our model on the Xi′an taxi GPS trajectories dataset. Experiments show that our proposed network can effectively predict the temporal evolution of vehicle emissions.

Keywords: Vehicle emission prediction, spatiotemporal gragh convolution, GPS trajectories, motor vehicle emission simulator (MOVES) model, feature sharing.

1 Introduction

Environmental pollution from traffic is becoming an important issue that people are concerned about. Urban traffic congestion and dense traffic flow have exacerbated environmental problems, and vehicle emissions have become one of the main sources of urban air pollution[1]. The pollutants emitted by vehicles contain carbon monoxide, nitrogen oxides, and particulate matter, which are the main causes of smog and photochemical smog pollution[2]. Therefore, the urban transportation system needs to establish an effective environmental monitoring and early warning system. The key issue is the accurate

prediction of the evolution of traffic emissions. The trend of vehicle pollution emissions is mainly affected by driving conditions, such as changes in vehicle speed, acceleration, and traffic volume. Predicting the emission of motor vehicles in the future means predicting the temporal and spatial trends of multiple related traffic condition variables. Therefore, the accurate and efficient prediction of trends of multiple related traffic condition variables can provide scientific estimations for predicting the evolution of urban vehicle emissions.

In recent years, more and more researchers have focused their research on vehicle emissions and urban air pollution. The existing researches on the prediction of the temporal and spatial distribution of urban mobile source emissions can be roughly divided into two main methods: model-driven and data-driven. Based on the model-driven method, vehicle emission models are proposed to obtain an emission inventory, which is used to estimate or predict the total emission traffic of a specific area[3, 4],

within a specified time range (usually one year). Wang et al.[5] proposed a vehicle emission model based on mileage and emission factors to study vehicle emission trends in China. Emission models such as motor vehicle emission simulator (MOVES)[6], computer programme to calculate emissions from road transport (COPERT)[7, 8] and the international vehicle emission (IVE) model[9] have been developed and adjusted according to vehicle information databases (such as vehicle type and fuel type) in various locations. MOVES was developed by the US Environmental Protection Agency and used to estimate emissions from mobile sources on highways. It considers several mobile emissions processes, including exhaust during driving, brake wear, tire wear, and driving losses. COPERT is a commonly used emission model in Europe. The model uses lots of experimental data to determine the emission parameters of road transportation and obtain an emission inventory. The IVE model uses vehicle specific power (VSP) and engine stress (ES) as inputs to calculate emission factors. Recently, Jamshidnejad et al.[10] proposed a comprehensive framework that combines micro and macro emission models to estimate vehicle emissions. The emission inventory estimated by the model-driven methods can provide the macro-emissions of the city, but it cannot satisfy the short-term and fine-grained forecasting needs of the early warning mechanism in an urban environmental monitoring system.

Although the above methods have made great progress, it is still a challenging problem to predict vehicle emissions. The model-driven algorithms lack universality and ignore the influence of geographic and environmental factors on the distribution of traffic flow. Due to the limitation of accuracy, the above algorithm can only predict the total emissions of the entire city or region, and cannot predict emissions within the fine-scale range.

With the rapid development of traffic data collection and big data technology[11, 12], researchers have turned to data-driven methods, using mobile source pollution emission monitoring data and other urban multi-source data to study spatiotemporal prediction of vehicle emissions. The development of laser remote sensing technology[13–16] has enabled on-road monitoring equipments for the remote sensing of emissions to measure the instantaneous emissions of vehicles while driving. Xu et al.[17] proposed long short-term memory (LSTM) networks combined with automatic encoders to predict vehicle emissions based on data obtained from the remote sensing station. Xu et al.[18] proposed a deep spatiotemporal residual early-late fusion network with the semi-supervised geographical weighted regression to predict vehicle emissions in urban areas, using the sparse monitoring stations. The data collected in this way is not extensive enough, and the sparseness of the monitoring stations makes the granularity of the emission predictions insufficient.

With the establishment and improvement of the environmental monitoring system and intelligent transportation system, a large number of GPS sensors are installed on taxis. The trajectory data with the spatiotemporal properties are distributed on various streets of the city, and can reflect the traffic state in a small area or even on the street scale. GPS trajectories have time continuity and closeness, which can realize short-term estimation. More than that, the data covers all the streets of the city, which not only overcomes the sparseness of telemetry stations, but also enables fine-grained emission analysis.

Aslam et al.[19] verified that the traffic patterns reflected in the taxi trajectory data obtained through high-density sampling have roughly the same trend as the actual traffic patterns. Therefore, it is widely used in the research of urban traffic and traffic pollution. Nyhan et al.[20] inferred the spatial and temporal distribution of Singapore's vehicle emissions by taxi GPS trajectories and loop detector data. Shang et al.[21] used taxi trajectory data and urban road network information to infer vehicle emissions in Beijing. Nocera et al.[22] estimated the carbon emissions of road transport using incomplete traffic information collected by the flow estimator.

The existing research on using big data to estimate vehicle pollution emissions at any scale is mature. In most works, the current or past vehicle emission distribution is estimated with the aid of traffic conditions, and the analysis of spatiotemporal evolution of emissions is blank temporarily.

It is a prediction problem to obtain evolution of vehicle emissions by using historical GPS trajectory data. There are some challenges here:

1) The prediction of emission trends is complicated and requires high accuracy and calculation efficiency. Only considering a single feature cannot achieve accurate modeling and prediction of emissions. In order to improve the accuracy of vehicle emission prediction, multiple traffic attributes need to be considered in the vehicle emission model. For example, the traffic average speed and volume are used in the COPERT model to calculate the emissions of vehicles; the MOVES model additionally considers average acceleration. However, predicting multiple features separately will greatly increase the computation cost and reduce the development efficiency. Therefore, how to improve calculation efficiency in the case of predicting multiple traffic state-related features at the same time becomes an urgent problem to be solved.

2) How to effectively extract and integrate multiple attributes related to traffic emissions is also a challenging task. The attributes related to traffic emissions include traffic volume, average speed, acceleration, etc. The temporal and spatial variation trend of the attributes are complex nonlinear models and are hard to predict precisely. Besides, affected by the external environment and location factors, these attributes reflect different traffic conditions, but at the same time they are related to each other, which brings difficulties to the design of feature extraction and fusion mechanism.

In order to solve the above problems, we propose a joint framework based on the multi-channel spatiotemporal graph convolution network and the MOVES emission factor model to predict spatial and temporal distribution of traffic emissions with reasonable accuracy, using multi-source datasets including meteorological data and road network data.

Specifically, we first match the GPS trajectories to different spatial grids with a map matching algorithm, and extract traffic status that changes with time, which is the volume, average speed and acceleration of passing taxis in each grid. Then spatiotemporal feature spaces of taxi volume, speed and acceleration can be constructed by multi-channel spatiotemporal graph convolution networks. We will use the feature sharing mechanism to couple above features to predict traffic states in future periods. Finally, the MOVES emission model is used to calculate emission factors, combined with road network information to estimate the emission evaluation of vehicles. Through the comparison and visualization experiments on the Xi'an taxi trajectories dataset, the effectiveness of this method is proved. The main contributions of this paper are as follows:

1) In order to improve the accuracy of pollution prediction and ensure the computational efficiency of the model, we use a multi-channel mechanism to achieve simultaneous prediction of multiple attributes. At the same time, in order to balance the scale differences of different features, we introduce homoscedastic uncertainty to learn the weight of the loss of each channel.

2) In order to better construct the spatiotemporal dependence of each attribute, we use spatiotemporal graph convolutional network (STGCN) in each channel to extract spatiotemporal features layer by layer. In addition,

we introduce a feature sharing mechanism to model the selection tendency in the extraction of features between related attributes which helps to achieve better fusion and utilization of different attributes.

3) We evaluate the effectiveness of our method on the GPS trajectory dataset of taxis in Xi'an. The results show that the multi-channel mechanism shortens the training time by 17.72%, and under the premise of ensuring the prediction accuracy, the prediction accuracy of traffic volume and average speed attributes are respectively increased by 4.86% and 4.68%, proving the effectiveness of the feature sharing mechanism. In addition, through the prediction and analysis of pollution in different functional areas of the city, the distribution of pollution is basically consistent with the actual situation, so it can be considered that the prediction of vehicle emissions is effective.

The rest of this paper is organized as follows. System overview is written in Section 2. Section 3 describes the proposed urban traffic emission evolution prediction model. The details of the experimental setup and the results of the related experiments are written in Section 4. Finally, we conclude this paper in Section 5.

2 Definition

Definition 1. Trajectory. Let $\mathbf{P}$ denote a set of GPS trajectories at the t-th time interval, and $Tr : p_1 \to p_2 \to \cdots \to p_n$ is a trajectory in $\mathbf{P}$, where p has a geospatial coordinate set g and a timestamp τ, $p = (\tau, g)$.

Definition 2. Node. Dividing the city into $N = I \times J$ grid based on the longitude and latitude, denoted by $V = \{r_1, r_2, \cdots, r_N\}$, each of which represents a
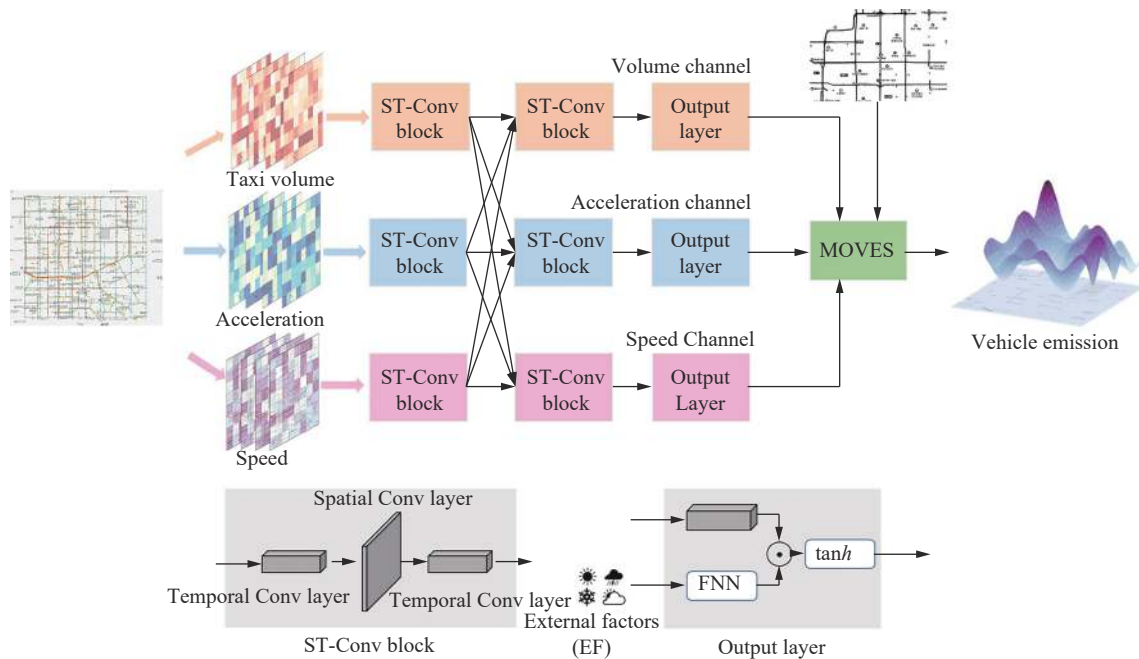

Fig. 1   Architecture of multi-channel spatiotemporal graph convolutional networks and MOVES model

spatial node.

3 Methodology

In this section, we propose the prediction model for evolution of traffic emissions based on a joint training framework of map-matching and multi-channel spatiotemporal graph convolution network, as shown in Fig. 1. Specifically, we first match the GPS trajectory to corresponding spatial grids using the map-matching algorithm, and taxi volume, average speed and acceleration in each grid are calculated. Then, we adopt a multi-channel spatiotemporal graph convolution network to construct the feature spaces of volume, speed and acceleration respectively, and use the feature sharing mechanism to couple the above features to predict the traffic states in spatial grids in the next time interval. Finally, we estimate the emission of urban vehicles by the MOVES model, using road network and urban vehicle statistical information.

3.1 Traffic states construction

The GPS trajectories received by the vehicle are projected onto the road network using the map-matching algorithm[23]. After matching, each point of the trajectories is mapped onto the corresponding road segment. Given two trajectory points p_1 and p_2, the speed and acceleration of p_1 can be calculated:

$$v_1 = \frac{dist\,(p_1.g, p_2.g)}{|p_2.\tau - p_1.\tau|} \tag{1}$$

$$a_1 = \frac{v_1 - v_2}{|p_2.\tau - p_1.\tau|} \tag{2}$$

where $dist\,(\cdot)$ is the function that calculates the distance of road network between two points. Likewise, we can calculate the average speed and average acceleration of grid r_i at the interval t as follows:

$$\bar{v}_t^i = \frac{1}{|\mathbf{P}^i| \times |Tr^i|} \sum_{Tr^i \in \mathbf{P}^i} \sum_{p_k \in Tr^i} v_k \tag{3}$$

$$\bar{a}_t^i = \frac{1}{|\mathbf{P}^i| \times |Tr^i|} \sum_{Tr^i \in \mathbf{P}^i} \sum_{p_k \in Tr^i} a_k \tag{4}$$

where $|\cdot|$ denotes the cardinality of the set. $Tr^i = \{p_k \mid p_k.g \in r_i\}$ and $\mathbf{P}^i = \{Tr \mid Tr = Tr^i\}$ denote a set of trajectory points and a set of trajectories in grid r_i. In addition, GPS trajectory data can be used to find the traffic volume of a certain area in a time interval. The taxi volume of grid r_i at the interval t is defined as

$$f_t^i = \sum_{Tr \in \mathbf{P}} |\{k \geq 1 \mid p_k.g \in r_i\}|. \tag{5}$$

Therefore, three matrices representing time-varying

traffic conditions within the grid area can be extracted from the historical trajectory. Firstly, we define the set of time intervals as $\mathcal{T} = \{t_1, t_2, \cdots, t_T\}$, and traffic status including average speed, acceleration and taxi volume in all N regions can be denoted as three tensors S, A, $F \in \mathbf{R}^{T \times N}$, where $S[i,t] = \bar{v}_t^i$, $A[i,t] = \bar{a}_t^i$ and $F[i,t] = f_t^i$. The area within Xi′an Second Ring Road is divided into 8×8 disjoint grids. Each row of the three matrices represents a time interval, and each column represents a grid unit. Each element represents the average speed, acceleration, and taxi volume of all vehicles passing through the grid area in certain time interval, which can reflect the traffic condition of the area. As show in Fig. 2, when the trajectories in the grid are dense, i.e., the taxi volume is large, the average speed of vehicles is lower than that in the sparse trajectory area. Similarly, the average acceleration of vehicles is also lower than the grid with fewer vehicles.

3.2 Multi-channel spatiotemporal graph convolution network

3.2.1 Graph construction

In this work, we define the road network as a set of time-varying spatial graphs $\mathcal{G}$. In graph $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}, W)$ at the t-th time interval, $\mathcal{V}_t$ is a set of vertices corresponding to the traffic status $S_t, A_t, F_t \in \mathbf{R}^N$ in the above-mentioned nodes; $\mathcal{E}$ is the set of edges representing the connectedness between nodes, while $W \in \mathbf{R}^{N \times N}$ denotes the weighted adjacency matrix of $\mathcal{G}$.

3.2.2 Multi-channel STGCN

The multi-channel spatiotemporal graph convolutional network is a model that we proposed based on STGCN[24], which can predict the three features of taxi volume, average speed and acceleration in the traffic network simultaneously, and guarantee reasonable accuracy and scale. The network structure is shown in Fig. 1. A novel multi-channel feature sharing mechanism is proposed in our model. The network constructs the spatiotemporal feature space of volume, speed and acceleration separately, and interactively encapsulates them into the spatiotemporal convolution module of other channels. Three channels eavesdrop to feature information related to themselves through a feature sharing mechanism.

In this paper, spatiotemporal graph convolutional networks[24] are used to capture the dynamic spatial and temporal correlations on traffic networks. The network includes several spatiotemporal convolutional blocks, which are a combination of graph convolutional layers[25] and temporal convolutional layers to model spatial and temporal correlations.

Regarding the graph convolution layer, we mainly consider spectral convolution on arbitrary graphs[25]. Since it is difficult to express meaningful conversion operators in the node domain, the spectral representation of the convolution operator is given on the graph[25], de-
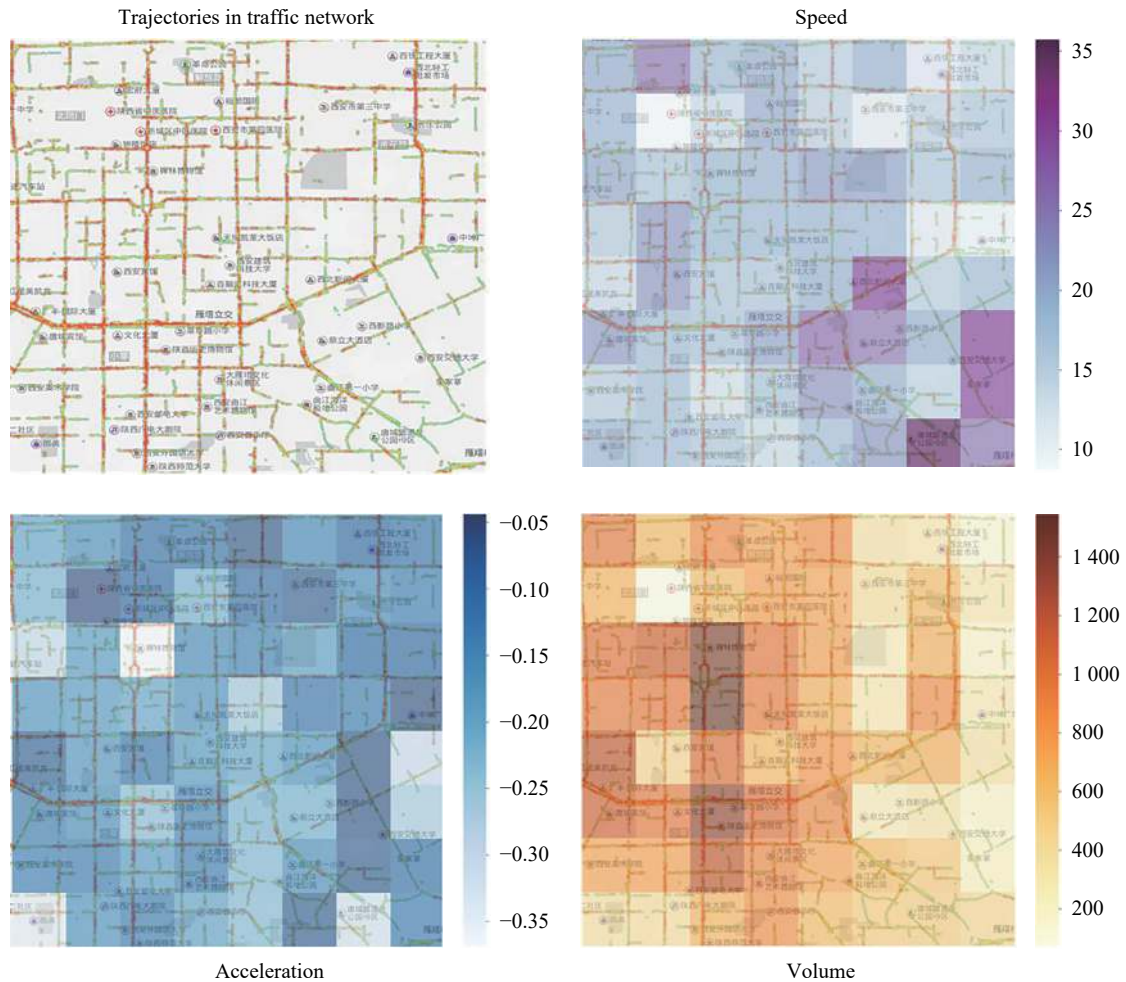
Trajectories in traffic network

Speed

Acceleration

Volume

Fig. 2   Visualization of GPS trajectories, taxi volume, speed, acceleration

noted as $*_{\mathcal{G}}$. According to above definition, node feature vector $\boldsymbol{X} \in \mathbf{R}^{N \times T}$ with a filter $g_w = \mathrm{diag}(\boldsymbol{w})$ parameterized by $\boldsymbol{w} \in \mathbf{R}^N$ in the Fourier domain is

$$g_w *_{\mathcal{G}} \boldsymbol{X} = g_w(\boldsymbol{L})\boldsymbol{X} = g_w\left(\boldsymbol{U\Lambda U}^{\mathrm{T}}\right)\boldsymbol{X} \qquad (6)$$

where $\boldsymbol{U} \in \mathbf{R}^{N \times N}$ is the eigenvector matrix and $\boldsymbol{\Lambda} \in \mathbf{R}^{N \times N}$ is the diagonal matrix of the eigenvalues of the normalized graph Laplacian L.

$$\boldsymbol{L} = \boldsymbol{I}_N - \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}} = \boldsymbol{U\Lambda U}^T \in \mathbf{R}^{N \times N} \qquad (7)$$

where $\boldsymbol{I}_N$ represents the identity matrix and $\boldsymbol{D} \in \mathbf{R}^{N \times N}$ represents the diagonal degree matrix with $\boldsymbol{D}_{ii} = \sum_j \boldsymbol{A}_{ij}$. We define g_w as the eigenvalue function of L. However, since the complexity of multiplying by $\boldsymbol{U}$ is $O\left(N^2\right)$, the calculation is huge. In order to solve the above problem, Defferrard et al.[25] used Chebyshev polynomial expansion to obtain effective approximation,

$$g_w(\boldsymbol{L})\boldsymbol{X} \approx \sum_{k=0}^{K-1} \boldsymbol{w}_k\left(\boldsymbol{L}^k\right)\boldsymbol{X} = \sum_{k=0}^{K-1} \boldsymbol{w}_k' T_k(\tilde{\boldsymbol{L}})\boldsymbol{X} \qquad (8)$$

where $T_k(\tilde{L})$ is the k-th order Chebyshev polynomial

evaluated on the scaled Laplace operator $\tilde{\boldsymbol{L}} = \dfrac{2}{\lambda_{\max}}\boldsymbol{L} - \boldsymbol{I}_N$, $\lambda_{\max}$ is the largest eigenvalue of $\boldsymbol{L}$, and $\boldsymbol{w}' \in \mathbf{R}^K$ is the vector of Chebyshev coefficients. Details about Chebyshev polynomials approximation can be found in [25, 26].

As for temporal convolutional layers, it contains a one-dimensional convolution, the width of the kernel is K_t, and then gated linear units (GLU)[27] is connected as the activation unit. The input of the temporal convolution for each node in graph $\mathcal{G}$ can be regarded as a sequence of length M, with C_i channels, denoted as $Y \in \mathbf{R}^{M \times C_i}$. The convolution kernel $\Gamma \in \mathbf{R}^{\mathrm{Kt} \times C_i \times 2C_o}$ is designed to map Y to a single output. Therefore, temporal convolutional layers can be defined as

$$\Gamma *_{\mathcal{T}} Y = P \odot \sigma(Q) \in \mathbf{R}^{(M-K_t+1) \times C_o} \qquad (9)$$

where P and Q are the input gates of GLU, $\odot$ donates the Hadamard product. The sigmoid activation function $\sigma(Q)$ controls which input P of the current state is relevant to discover the composition structure and dynamic variance in the time series. A non-linear activation function can perform deep mining on inputs

🍎 Springer

field by stacking temporal convolutional layers. In addition, when stacking temporal convolutional layers, residual connections are realized. Similarly, the same convolution kernel is used on each node, and the temporal convolution can be generalized to 3D variables, denoted as $\Gamma *_\mathcal{T} \mathcal{Y}$ with $\mathcal{Y} \in \mathbf{R}^{M \times N \times C_i}$.

In order to fuse features from spatial and temporal domains at the same time, Yu et al.[24] constructed a spatiotemporal convolutional block (ST-Conv block) based on the bottleneck strategy, including two temporal convolutional layers, respectively in the upper and lower two layers, and a spatial convolution layer in the middle. When the input of the block l is the characteristic matrix $\boldsymbol{x}^l \in \mathbf{R}^{M \times n \times C^l}$, then the output $\boldsymbol{x}^{l+1} \in \mathbf{R}^{(M-2(K_t-1)) \times n \times C^{l+1}}$ is calculated:

$$\boldsymbol{x}^{l+1} = \Gamma_1^l *_\mathcal{T} \mathrm{ReLU}\left(\Theta^l *_\mathcal{G} \left(\Gamma_0^l *_T \boldsymbol{x}^l\right)\right) \qquad (10)$$

where Γ_0^l, Γ_1^l are the upper and lower temporal convolutional layers parameters of block l, Θ^l is the graph convolution spectrum kernel, and $\mathrm{ReLU}(\cdot)$ represents the ReLU activation function. After stacking two ST-Conv blocks, a temporal convolutional layer and a fully-connected layer are used as the final output layer.

We design three parallel channels, namely volume channel, speed channel, and acceleration channel to extract the temporal and spatial dependence feature of multiple attributes such as traffic flow, speed and acceleration respectively, as shown in Fig. 1. Each channel in multi-channel spatiotemporal graph convolutional network (MC-STGCN) is a ST-subnetwork, which is composed of two ST-Conv blocks and an output layer. Input of the three sub-networks is time-ordered sequence of traffic attribute graphs. Under the actions of parallel networks, feature extraction processes of attributes are independent of each other and proceed simultaneously, which improves computational efficiency of our task.

In order to fuse features from spatial and temporal domains at the same time, Yu et al.[24] constructed the spatiotemporal convolutional block (ST-Conv block) based on the bottleneck strategy, including two temporal convolutional layers, respectively in the upper and lower, and a spatial convolution layer in the middle. When the input of the block l is the characteristic matrix $\boldsymbol{s}^l \in \mathbf{R}^{M \times n \times C^l}$ of the traffic speed, the characteristic matrix of its acceleration is also packaged as the input of the block l, then the output $\boldsymbol{s}^{l+1} \in \mathbf{R}^{(M-2(K_t-1)) \times n \times C^{l+1}}$ is calculated:

$$\boldsymbol{s}^{l+1} = \Gamma_1^l *_\mathcal{T} \mathrm{ReLU}\left(\Theta^l *_\mathcal{G} \left(\Gamma_0^l *_\mathcal{T} \left[\boldsymbol{s}^l, \boldsymbol{a}^l, \boldsymbol{f}^l\right]\right)\right) \qquad (11)$$

$$\boldsymbol{a}^{l+1} = \Gamma_1^l *_\mathcal{T} \mathrm{ReLU}\left(\Theta^l *_\mathcal{G} \left(\Gamma_0^l *_\mathcal{T} \left[\boldsymbol{a}^l, \boldsymbol{s}^l, \boldsymbol{f}^l\right]\right)\right) \qquad (12)$$

$$\boldsymbol{f}^{l+1} = \Gamma_1^l *_\mathcal{T} \mathrm{ReLU}\left(\Theta^l *_\mathcal{G} \left(\Gamma_0^l *_\mathcal{T} \left[\boldsymbol{f}^l, \boldsymbol{s}^l, \boldsymbol{a}^l\right]\right)\right) \qquad (13)$$

where Γ_0^l, Γ_1^l are the upper and lower temporal convolutional layers parameters of block l, Θ^l is the graph convolution spectrum kernel, and $\mathrm{ReLU}(\cdot)$ represents the ReLU activation function. After stacking two ST-Conv blocks, a temporal convolutional layer and a fully connected layer are used as the final output layer.

3.2.3 Feature sharing mechanism

For traffic state prediction models including taxi volume, traffic speed and acceleration, they have relatively close temporal and spatial characteristics, such as spatial correlation, temporal periodicity and dependence, etc. Due to the similarity between channels, feature sharing can provide more information about spatiotemporal characterization for each task, thereby assisting ST-sub-networks to extract more accurate feature representations.

In this article, the output of the first ST-Conv block in each channel will be used as a shared feature and become the input feature of the second ST-Conv block in other channels. Since the feature representations of different channels have different scales and statistical features, the network will prefer features with larger values and ignore other feature information. Therefore, we first standardize the shared features and then concatenate them into high-dimensional feature vectors. The input of the second ST-Conv block of speed channel is

$$s^2 = \mathrm{Concat}(\mathrm{Norm}(s^1), \mathrm{Norm}(a^1), \mathrm{Norm}(f^1)) \in \mathbf{R}^{3M \times n \times C^1} \qquad (14)$$

where s^1, a^1, $f^1 \in \mathbf{R}^{M \times n \times C^1}$ are outputs of the first ST-Conv block in three channels, and $\mathrm{Norm}(\cdot)$ represents standardized operation, which transforms features into a representation with same mean and variance. $\mathrm{Concat}(\cdot)$ is to concatenate matrix according to a certain dimension. Similarly, the inputs of second ST-Conv block of acceleration channel and volume channel are:

$$a^2 = \mathrm{Concat}(\mathrm{Norm}(a^1), \mathrm{Norm}(s^1), \mathrm{Norm}(f^1)) \qquad (15)$$

$$f^2 = \mathrm{Concat}(\mathrm{Norm}(f^1), \mathrm{Norm}(s^1), \mathrm{Norm}(a^1)). \qquad (16)$$

3.2.4 External factor fusion

External factors like weather can affect urban traffic and road conditions. For example, a heavy rain may congest the streets. External factors are like switches, and if they do, the road conditions will change dramatically. Based on this conclusion, Zhang et al.[28] develop a gating-mechanism-based fusion, which can obtain the corresponding external features expressed as $E_t \in \mathbf{R}^i$ at time t in the network, as shown in Fig. 1. Formally, we can get

the following gating value:

$$F_t = \sigma\left(\boldsymbol{W}_e \cdot E_t + b_e\right) \qquad (17)$$

where $\boldsymbol{W}_e \in \mathbf{R}^{l_e \times 1}$ and $b_e \in \mathbf{R}^1$ are learnable parameters. $F_t \in \mathbf{R}^0$ is the output value of the gate. $\sigma(\cdot)$ is the sigmoid function and "·" is dot product of two matrices. Then a product fusion based on the gating mechanism is embedded in the output layer of each ST-subnetwork for speed $\boldsymbol{s}_{in}$, acceleration $\boldsymbol{a}_{in}$ and volume $\boldsymbol{f}_{in}$:

$$\boldsymbol{s}_{out} = \tanh\left(F_{t+1} \odot \left(\Gamma_s^o *_\mathcal{T} \boldsymbol{s}_{in}\right)\right) \qquad (18)$$

$$\boldsymbol{a}_{out} = \tanh\left(F_{t+1} \odot \left(\Gamma_a^o *_\mathcal{T} \boldsymbol{a}_{in}\right)\right) \qquad (19)$$

$$\boldsymbol{f}_{out} = \tanh\left(F_{t+1} \odot \left(\Gamma_f^o *_\mathcal{T} \boldsymbol{f}_{in}\right)\right) \qquad (20)$$

where tanh is a hyperbolic tangent function, which ensures that the output value is between -1 and 1. Γ is the parameter of the temporal convolutional layer, and F_{t+1} represents the weather forecast of environmental features at future moments. According to this, the predicted values $\hat{s}_{t+1}, \hat{a}_{t+1}, \hat{f}_{t+1}$ at time $t+1$ are as follows:

$$\begin{aligned} \hat{s}_{t+1} &= \boldsymbol{W}_s * \boldsymbol{s}_{out} + b_s \\ \hat{a}_{t+1} &= \boldsymbol{W}_a * \boldsymbol{a}_{out} + b_a \\ \hat{f}_{t+1} &= \boldsymbol{W}_f * \boldsymbol{f}_{out} + b_f. \end{aligned} \qquad (21)$$

3.2.5 Loss function design for MC-STGCN

$\boldsymbol{W}_\Phi$ is the set of all learnable parameters of predicted speed models of MC-STGCN. Our goal is to learn them by minimizing the loss function between the predicted value $\hat{S}$ and the true value S:

$$L\left(\hat{s}; \boldsymbol{W}_\Phi\right) = \sum_t \|\hat{s}\left(s_{t-M+1}, \cdots, s_t, \boldsymbol{W}_\Phi\right) - s_{t+1}\|^2. \quad (22)$$

Similarly, $\boldsymbol{W}_\Theta$ and $\boldsymbol{W}_\Psi$ is a set of all learnable parameters of predicted acceleration models and volume models. For the square loss function,

$$L\left(\hat{a}; \boldsymbol{W}_\Theta\right) = \sum_t \|\hat{a}\left(a_{t-M+1}, \cdots, a_t, \boldsymbol{W}_\Theta\right) - a_{t+1}\|^2 \qquad (23)$$

$$L\left(\hat{f}; \boldsymbol{W}_\Psi\right) = \sum_t \left\|\hat{f}\left(f_{t-M+1}, \cdots, f_t, \boldsymbol{W}_\Psi\right) - f_{t+1}\right\|^2 \qquad (24)$$

where $s_{t+1}, a_{t+1}, f_{t+1}$ are ground truths, $\hat{s}(\cdot), \hat{a}(\cdot), \hat{f}(\cdot)$ represent predicted values of model.

The performance of the multi-channel learning model depends on the loss weight between channels. Manually adjusting the weights is time-consuming and labor-intensive. In order to better optimize our proposed multi-channel network, we use the strategy proposed in [29] to balance the three channels. Then, the network loss function

formula is as follows:

$$\begin{aligned} L\left(\boldsymbol{W}_\Phi, \boldsymbol{W}_\Theta, \boldsymbol{W}_\Psi, \sigma_s, \sigma_a, \sigma_f\right) = {} & \frac{1}{2\sigma_s^2} L\left(\hat{s}; \boldsymbol{W}_\Phi\right) + \\ & \frac{1}{2\sigma_a^2} L\left(\hat{a}; \boldsymbol{W}_\Theta\right) + \frac{1}{2\sigma_f^2} L\left(\hat{f}; \boldsymbol{W}_\Psi\right) + \log \sigma_s^2 \sigma_a^2 \sigma_f^2 \end{aligned}$$
$$(25)$$

where σ_s, σ_a, σ_f are balancing weights of three channels, which can be optimized as parameters in training. The loss function automatically learns the weighted hyperparameters in the loss function through the homoscedastic uncertainty, so that the loss function of each task has a similar scale.

3.3 Emission model

Traffic speed, acceleration and volume information can be further used to estimate vehicle emissions in the road network. Different models can be used in environmental science. These models can quantify the relationship between emissions and speed and other factors based on large amounts of data. The MOVES was developed by the US Environmental Protection Agency (EPA) and is capable of calculating vehicles pollutant emissions at different scales. The reason why we use this model is that MOVES can more meticulously describe the working conditions and emission levels of vehicles. In the MOVES model, it mainly calculates the distribution of operating conditions of vehicles, and combines speed, acceleration and vehicle specific power (VSP).

$$VSP_{v,t} = \frac{Av_t + Bv_t^2 + Cv_t^3 + Mv_t a_t + mgv_t \sin\theta}{M} \quad (26)$$

where:

A: Rolling resistance coefficient kW/(m/s)

B: Rotational resistance coefficient kW/(m^2/s^2)

C: Aerodynamic drag coefficient kW/(m^3/s^3)

M: Vehicle mass, tonne

v_t: Speed at time t (m/s)

a_t: Acceleration at time t (m/s^2)

θ: Slope

g: Acceleration of gravity, 9.8 m/s^2

For light vehicles, the calculation formula (22) is simplified to (23)[30],

$$VSP = v(1.1a + 0.132) + 0.000\,302v^3. \qquad (27)$$

A simplified MOVES model is proposed in [31], which divides the operating mode of vehicles into 23 types, corresponding to different default average emission rates (AER), as shown in Table 1. The values are given in Table 1 to calculate different kinds of emissions, which applies for Euro III passenger vehicles. Although diversity of vehicles will influence accuracy, the results are still statistically useful as we sample the most represent-

Table 1   Operating mode bin average emission rates in the MOVES model for selected pollutants

Operating mode description		Average emission rate (g/h)			
		CO_2	NOx	CO	HC
Braking		3 529	0.23	5.14	0.19
Idling		3 265	0.10	0.89	0.05
$1 \leq$ speed < 25	VSP < 0	5 134	0.34	17.69	0.13
	$0 \leq$ VSP < 3	7 089	0.52	28.88	0.10
	$3 \leq$ VSP < 6	9 852	1.22	26.62	0.19
	$6 \leq$ VSP < 9	12 449	2.15	38.20	0.26
	$9 \leq$ VSP < 12	14 845	3.81	55.39	0.36
	$12 \leq$ VSP	17 930	7.94	93.47	0.58
$25 \leq$ speed < 50	VSP < 0	6 985	0.67	23.05	0.20
	$0 \leq$ VSP < 3	7 950	1.09	30.55	0.18
	$3 \leq$ VSP < 6	9 683	1.65	39.28	0.20
	$6 \leq$ VSP < 9	12 423	2.79	57.42	0.38
	$9 \leq$ VSP < 12	16 578	3.91	65.17	0.37
	$12 \leq$ VSP < 18	21 855	6.16	97.87	0.59
	$18 \leq$ VSP < 24	29 459	13.54	239.24	3.84
	$24 \leq$ VSP < 30	40 359	23.78	506.67	6.81
	$30 \leq$ VSP	50 682	31.29	1 779.51	11.25
$50 \leq$ speed	VSP < 6	9 951	1.44	17.31	0.19
	$6 \leq$ VSP < 12	15 956	3.96	29.56	0.27
	$12 \leq$ VSP < 18	20 786	5.54	43.51	0.34
	$18 \leq$ VSP < 24	27 104	11.50	219.28	2.59
	$24 \leq$ VSP < 30	36 102	17.12	231.37	3.76
	$30 \leq$ VSP	46 021	21.56	679.99	4.92

ative vehicles in calculation.

The dataset applied in this paper is taxi GPS trajectories collected in the Second Ring of Xi′an city, and speed of vehicles in urban area is generally limited between 40 km/h and 60 km/h. Moreover, according to statistical characteristics of speed, that is the average is 30.26 km/h and standard deviation is 9.66 km/h. Therefore, the situation in Table 1 satisfies emission estimation under almost all driving states.

Therefore, the emission factor (EF) which is the amount of pollutants (g/km) emitted by each vehicle per kilometer is calculated,

$$EF = AER \times v. \tag{28}$$

The total emissions in particular area is

$$E = EF \times f \times L/1000 \tag{29}$$

where L is the road length, and f is traffic volume. Traffic volume is estimated by the ratio of the predicted taxi volume and the total number of urban vehicles. For example, in 2018, there were 3.24 million vehicles in Xi′an, and 86 267 taxis were collected in the data set. Therefore, the data sample accounts for 2.67% of the total vehicles in urban. The traffic volume in each grid is estimated by dividing taxi volume in the grid by the ratio.

4 Experiment

4.1 Dataset description

In this paper, the dataset we used is shown in Table 2, the details are as follows.

We use a large-scale online taxi GPS dataset collected by Didi Chuxing, which is an online car-hailing company in China. The data source is https://gaia.didichuxing.com. The dataset contains taxi GPS trajectories from October 1, 2018 to November 29, 2018 in Xi′an. The

Table 2   Datasets discription

Dataset	Taxi in Xi′an
Data type	Taxi GPS
Location	Xi′an
Time span	October 1, 2018 − November 29, 2018
Time interval	1 h
Grid map size	(8, 8)
Trajectory data	
Average sampling rate	2−4 s
Taxis	86 267
Available time interval	1 440
Road Network	
Spatial range	8 km × 8 km
Total length	514 km
External factors	
Temperature	[− 6 °C, 26 °C]
Wind speed	[0, 32 km/h]

dataset covers urban area of 8 km × 8 km within the Second Ring Road in Xi′an, and 46.6 million GPS points of 20 000+ taxis approximately every day. The dataset includes the geographic location of each vehicle and the corresponding time stamp, which is collected every 2−4 s.

The road network within the Second Ring Road of Xi′an includes a spatial range of 8 km × 8 km, with a total length of 514 thousands meters. Urban roads are divided into 4 levels, which are expressways, main roads, secondary roads and branch roads.

4.2 Data preprocessing

We divide the Second Ring road area of Xi′an into 8 × 8 grids, and the size of each area is about 1 km × 1 km. 1 hour is set as the length of the time interval, then nodes in graph contain 24 data points per day. We use the Z-score method to convert traffic speed and acceleration to a scale with mean 0 and variance 1. In experiment, the data from October 1, 2018 to November 17, 2018 (48 days) was used for training, and the data from November 18, 2018 to November 29, 2018 (12 days) was used as the validation set and the test set. When testing the prediction results, we use the first 12 time intervals to predict the value of the next time interval.

The adjacency matrix is calculated based on the distance between grids in the road network. In the paper, we use dynamic time warping[32] to calculate the similarity distance between node (grid) i and node (grid) j, $d_{ij} = DTW(i, j)$. The weighted adjacency matrix W is as follows:
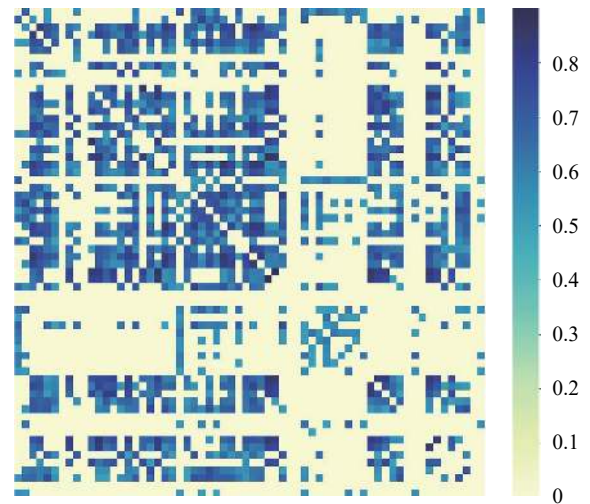

Fig. 3   Adjacency matrix

$$w_{ij} = \begin{cases} \exp\left(-\dfrac{d_{ij}^2}{\sigma^2}\right), & i \neq j \text{ and } \exp\left(-\dfrac{d_{ij}^2}{\sigma^2}\right) \geq \varepsilon \\ 0, & \text{otherwise.} \end{cases}$$

(30)

where w_{ij} is weight of the edge, which is decided by d_{ij}. And σ^2 and ε are thresholds that control distribution and sparsity of W, with 10 and 0.5, respectively. The visualization of $D = [d_{ij}]$ and W is shown in Fig. 3.

4.3 Experimental settings

We set hyperparameters of the network based on the performance on validation dataset. In our model, graph convolutional layers of the first and second ST-Conv block use 64 and 128 convolution kernals, respectively. All temporal convolutional layers use 32 convolution kernals, and adjust the temporal span of the data by controlling the step size of temporal convolution. During the training phase, the learning rate is 0.001 and the batch size is 16. In experiments, the multi-channel spatiotemporal network performance is evaluated by two common metrics: mean absolute error (MAE) and root mean square error (RMSE)

$$MAE = \frac{1}{n} \sum_i |y_i - \hat{y}_i|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_i (y_i - \hat{y}_i)^2}.$$

(31)

We compare MC-STGCN with widely used time series regression models, including:

1) HA: Historical average[33].
2) Static.
3) Var: Vector auto-regressive[34].
4) FNN: Feed-forward neural network[35].
5) LSTM: Long short-term memory network[36].

Table 3　Prediction results of MC-STGCN model and other baseline methods on Xi′an taxi datasets

Model	Taxi in Xi′an (1 hour) volume		Taxi in Xi′an (1 hour) speed		Taxi in Xi′an (1 hour) acceleration	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
HA	30.551	61.090	2.099	3.240	0.113	0.230
Static	96.496	148.815	3.256	4.753	0.142	0.299
Var	55.056	76.736	2.535	3.685	0.146	0.257
FNN	96.726	148.009	5.930	7.629	**0.105**	**0.214**
LSTM	63.944	121.760	2.946	4.972	0.114	**0.225**
GRU	65.924	122.699	3.032	5.009	0.108	0.225
STGCN	28.962	42.046	2.052	3.096	0.107	0.215
MC-STGCN	**27.553**	**39.249**	**1.956**	**3.005**	0.106	0.217
MC-STGCN (no-FS)	29.462	45.159	2.182	3.126	0.107	0.216

6) GRU: Gated recurrent units[37].

7) STGCN: Spatiotemporal graph convolutional networks[24].

All experiments are compiled and tested on a Linux cluster (CPU: Intel(R) Xeon(R) CPU E5-2 680 v3 @ 2.50 GHz, GPU: NVIDIA GeForce GTX 2 080).

4.4 Experiment results

We compare MC-STGCN with 7 baseline methods in the Taxi in Xi′an dataset. Table 3 shows the prediction performance results for the next hour. According to the evaluation indicators, we have achieved reasonably excellent performance in traffic volume, speed, and acceleration. As we can see, the prediction results of some general time series analysis methods (HA, static, FNN, LSTM, GRU) are usually not ideal, which shows that they only consider the temporal dependencies of features, and ignore spatial correlation. Therefore, these methods have limited ability to model nonlinear and complex traffic data. For Var, it further considers the spatial correlation between features. As a result, it achieves better performance. However, it fails to capture complex nonlinear temporal dependencies and dynamic spatial correlation. In contrast, STGCN is superior to other methods, indicating that it can effectively capture the dynamic changes of traffic data.

The MC-STGCN model we proposed is significantly better than the single channel result in the traffic speed feature, and slightly better than the single channel result in the traffic acceleration. This proves that the multi-channel feature sharing mechanism has advantages in assisting the network to extract spatiotemporal features.

Moreover, we designed an ablation experiment on feature sharing machanisms, which removes concatenation of feature vectors after the first ST-Conv block. The experiment results are in Table 3, in which MC-STGCN (no-FS) line is the result of a model without a feature sharing mechanism. Simply using multi-channel models to train three traffic attributes in parallel has a larger deviation than single-task STGCN. This is because the loss function of MC-STGCN contains errors of all attributes,
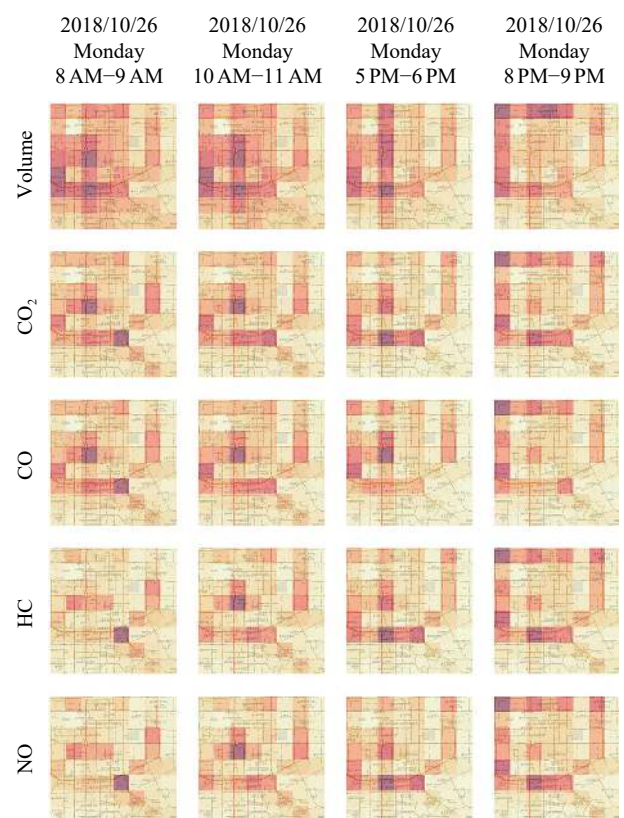

Fig. 4　Traffic volume, CO_2, CO, HC and NO on Monday

which makes it difficult to achieve the optimization effect of individual training. The introduction of feature sharing machanism makes up for this defect.

Fig. 4 visualizes the prediction of traffic volume, CO_2, CO, HC and NO in the Second Ring Road of Xi′an on a weekday (2018/11/26 Monday), respectively. As shown in the first line, in the time period from 8 AM to 9 AM, the traffic volume is larger than that from 10 AM to 11 AM. This is because of the morning peak, which is consistent with our common sense. Similarly, the volume distribution from 5 PM to 6 PM is different from volume distribution from 8 PM to 9 PM. When people get off work and
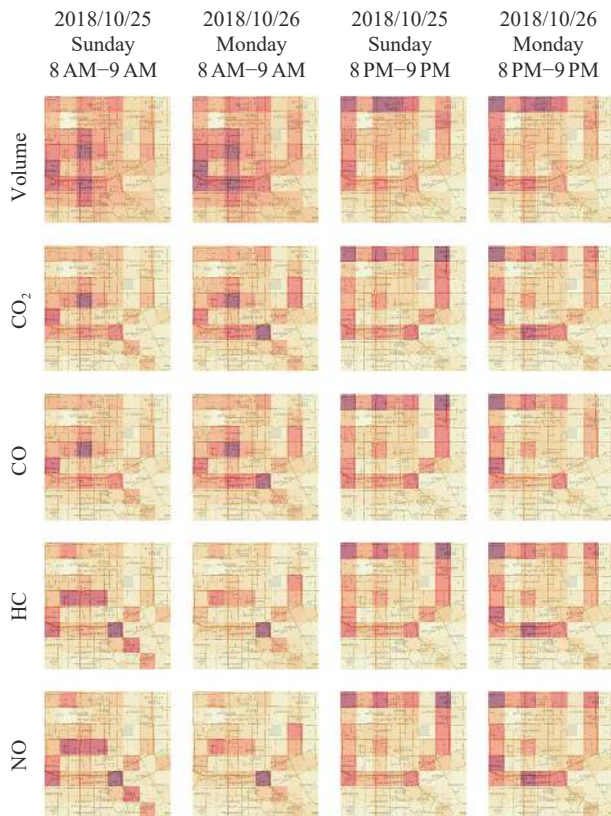
Fig. 5   Traffic volume, CO_2, CO, HC and NO on Sunday

Fig. 6   Comparison of traffic volume and vehicle emissions on weekdays and weekends

school, traffic volume from the work places in the city center to the residential areas. Therefore, the predicted vehicle emissions are mainly concentrated in dense roads in the city center during the morning and evening peak hours. After the evening peak, vehicle emissions increase in the suburban direction.

However, the distribution of traffic volume and vehicle emissions on weekends is different from that on weekdays. As shown in the first row of Fig. 5, the traffic volume during the period from 10 AM to 11 AM is significantly more than that from 8 AM to 9 AM. Judging by common sense, people do not have to go to work on weekends, usually go out late, and are more inclined to go to the entertainment area in the city center. Therefore, the vehicle emissions in the city center have increased during this period. Similarly, until 5 PM to 6 PM, the city center is still a gathering place for citizens, and vehicle emissions are still high. From 8 PM to 9 PM. in the evening, the traffic volume began to spread along the main urban roads, and the distribution of vehicle emissions also changed in the same trend.

Fig. 6 shows the comparison of traffic volume and vehicle emissions on weekdays and weekends. There is more traffic volume on weekdays from 8 AM to 9 AM than on weekends from 8 AM to 9 AM. In addition, the volume on weekdays is more concentrated than that on weekends, whether in the morning or at night. This is because people′s travel locations are more specific (work-

place or school) on weekdays, and the travel time is concentrated in the same period. On the contrary, on weekends, citizens′ travel locations, travel time, and the number of outgoing vehicles are scattered. Although the purpose of travel is different, the distribution of vehicles is roughly the same due to the concentration of workplaces and entertainment areas in the city center. Therefore, the pollutants emitted by vehicles in the morning and evening peaks are concentrated in the city center, and the vehicle emissions in the late night are higher in the suburbs. In addition, vehicle volume and pollutant emissions are distributed along urban roads. Large traffic volume in dense roads will cause congestion, and vehicles will emit more pollutants.
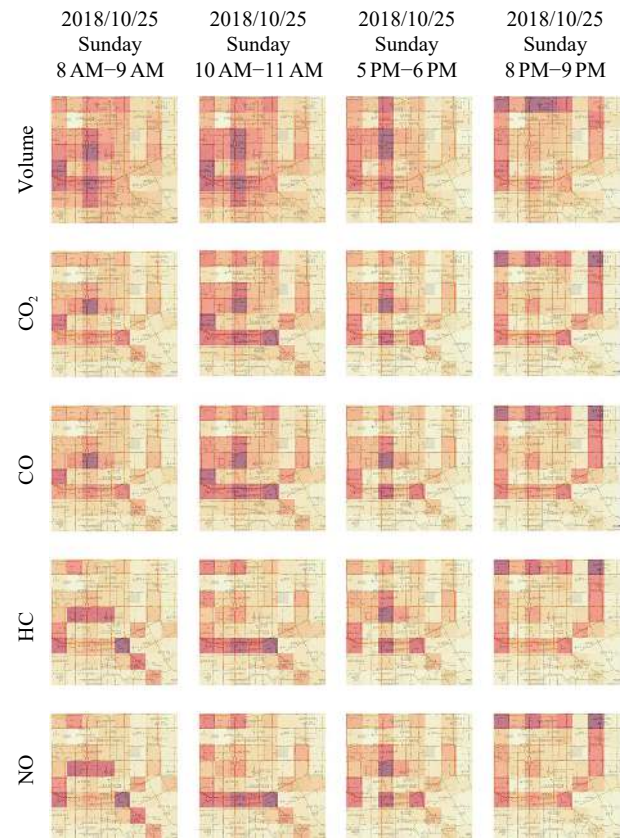
5   Conclusions

In this paper, we predict the evolution of vehicle emissions in urban road networks based on historical taxi GPS trajectories. The knowledge gained from our research can provide many valuable applications for social welfare, such as vehicle emission warnings, improving urban planning, and studying the sources of air pollution. Considering the efficiency and effectiveness, we solve this problem through a three-step method. Considering the efficiency and effectiveness, we solved this problem through a three-step method. We first map the trajectory data to

road networks and calculate the average traffic speed and acceleration of the area. And then, we use the multichannel STGCN network to predict the traffic status in future periods. Finally, the pollution emission distribution is calculated based on the predicted traffic status and the proportionally estimated traffic volume. We evaluate our method based on extensive experiments. The experiment uses GPS trajectories generated by more than 80 000 vehicles within two months. The results prove the effectiveness and rationality of our method. In the future, we will further improve the vehicle flow estimation algorithm to make the evolution of attributes related to traffic emissions more accurate.

Acknowledgments

Open Access

References

[1] World Health Organization. Urban Population Growth: Global Health Observatory, World Health Organization, Geneva, 2014.

[2] A. Aziz, I. U. Bajwa. Minimizing human health effects of urban air pollution through quantification and control of motor vehicular carbon monoxide (CO) in Lahore. *Environmental Monitoring and Assessment*, vol. 135, no. 1−3, pp. 459–464, 2007. DOI: 10.1007/s10661-007-9665-7.

[3] T. M. Butler, M. G. Lawrence, B. R. Gurjar, J. Van Aardenne, M. Schultz, J. Lelieveld. The representation of emissions from megacities in global emission inventories. *Atmospheric Environment*, vol. 42, no. 4, pp. 703–719, 2008. DOI: 10.1016/j.atmosenv.2007.09.060.

[4] Q. Y. Zhang, J. F. Xu, G. Wang, W. L. Tian, H. Jiang. Vehicle emission inventories projection based on dynamic emission factors: A case study of Hangzhou, China. *Atmospheric Environment*, vol. 42, no. 20, pp. 4989–5002, 2008. DOI: 10.1016/j.atmosenv.2008.02.010.

[5] H. K. Wang, L. X. Fu, Y. Zhou, X. Du, W. H. Ge. Trends in vehicular emissions in China's mega cities from 1995 to 2005. *Environmental Pollution*, vol. 158, no. 2, pp. 394–400, 2010. DOI: 10.1016/j.envpol.2009.09.002.

[6] J. Koupal, M. Cumberworth, H. Michaels, M. Beardsley, D. Brzezinski. Design and implementation of moves: EPA's new generation mobile source emission model. *Ann Arbor*, vol. 1001, Article number 48105, 2003.

[7] T. Nikoleris, G. Gupta, M. Kistler. Detailed estimation of fuel consumption and emissions during aircraft taxi operations at Dallas/fort worth international airport. *Transportation Research Part D: Transport and Environment*, vol. 16, no. 4, pp. 302–308, 2011. DOI: 10.1016/j.trd.2011.01.007.

[8] Z. Y. Xu, Y. Kang, Y. Cao, Z. R. Li. Deep amended COPERT model for regional vehicle emission prediction. *Science China Information Sciences*, vol. 64, no. 3, Article number 139202, 2021. DOI: 10.1007/s11432-018-9650-9.

[9] H. Guo, Q. Y. Zhang, Y. Shi, D. H. Wang. Evaluation of the international vehicle emission (IVE) model with on-road remote sensing measurements. *Journal of Environmental Sciences*, vol. 19, no. 7, pp. 818–826, 2007. DOI: 10.1016/S1001-0742(07)60137-5.

[10] A. Jamshidnejad, I. Papamichail, M. Papageorgiou, B. De Schutter. A mesoscopic integrated urban traffic flow-emission model. *Transportation Research Part C: Emerging Technologies*, vol. 75, pp. 45–83, 2017. DOI: 10.1016/j.trc.2016.11.024.

[11] J. R. Xue, J. W. Fang, P. Zhang. A survey of scene understanding by event reasoning in autonomous driving. *International Journal of Automation and Computing*, vol. 15, no. 3, pp. 249–266, 2018. DOI: 10.1007/s11633-018-1126-y.

[12] B. X. Wu, S. U. Ay, A. Abdel-Rahim. Pedestrian height estimation and 3D reconstruction using pixel-resolution mapping method without special patterns. *International Journal of Automation and Computing*, vol. 16, no. 4, pp. 449–461, 2019. DOI: 10.1007/s11633-019-1170-2.

[13] R. V. Martin. Satellite remote sensing of surface air quality. *Atmospheric Environment*, vol. 42, no. 34, pp. 7823–7843, 2008. DOI: 10.1016/j.atmosenv.2008.07.018.

[14] A. Van Donkelaar, R. V. Martin, R. J. Park. Estimating ground-level PM2. 5 using aerosol optical depth determined from satellite remote sensing. *Journal of Geophysical Research: Atmospheres*, vol. 111, no. D21, Article number D21201, 2006. DOI: 10.1029/2005JD006996.

[15] B. Zou, Q. Pu, M. Bilal, Q. L. Weng, L. Zhai, J. E. Nichol. High-resolution satellite mapping of fine particulates based on geographically weighted regression. *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 4, pp. 495–499, 2016. DOI: 10.1109/LGRS.2016.2520480.

[16] Z. Y. Xu, Y. Kang, Y. Cao. Emission stations location selection based on conditional measurement GAN data. *Neurocomputing*, vol. 388, pp. 170–180, 2020. DOI: 10.1016/j.neucom.2020.01.013.

[17] Z. Y. Xu, Y. Kang, Y. Cao, L. C. Yue. Residual autoencoder-LSTM for city region vehicle emission pollution pre-

diction. In *Proceedings of the 14th International Conference on Control and Automation*, IEEE, Anchorage, USA, pp. 811−816, 2018.

[18] Z. Y. Xu, Y. Cao, Y. Kang. Deep spatiotemporal residual early-late fusion network for city region vehicle emission pollution prediction. *Neurocomputing*, vol. 355, pp. 183–199, 2019. DOI: 10.1016/j.neucom.2019.04.040.

[19] J. Aslam, S. Lim, X. H. Pan, D. Rus. City-scale traffic estimation from a roving sensor network. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, ACM, New York, USA, pp. 141−154, 2012. DOI: 10.1145/2426656.2426671.

[20] M. Nyhan, S. Sobolevsky, C. G. Kang, P. Robinson, A. Corti, M. Szell, D. Streets, Z. F. Lu, R. Britter, S. R. H. Barrett, C. Ratti. Predicting vehicular emissions in high spatial resolution using pervasively measured transportation data and microscopic emissions model. *Atmospheric Environment*, vol. 140, pp. 352–363, 2016. DOI: 10.1016/j.atmosenv.2016.06.018.

[21] J. B. Shang, Y. Zheng, W. Z. Tong, E. Chang, Y. Yu. Inferring gas consumption and pollution emission of vehicles throughout a city. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, New York, USA, pp. 1027−1036, 2014. DOI: 10.1145/2623330.2623653.

[22] S. Nocera, C. Ruiz-Alarcón-Quintero, F. Cavallaro. Assessing carbon emissions from road transport through traffic flow estimators. *Transportation Research Part C: Emerging Technologies*, vol. 95, pp. 125–148, 2018. DOI: 10.1016/j.trc.2018.07.020.

[23] Y. Lou, C. Y. Zhang, Y. Zheng, X. Xie, W. Wang, Y. Huang. Map-matching for low-sampling-rate GPS trajectories. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, New York, USA, pp. 352−361, 2009. DOI: 10.1145/1653771.1653820.

[24] B. Yu, H. T. Yin, Z. X. Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, pp. 3634−3640, 2018. DOI: 10.24963/ijcai.2018/505.

[25] M. Defferrard, X. Bresson, P. Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, ACM, Red Hook, USA, pp. 3844−3852, 2016.

[26] D. K. Hammond, P. Vandergheynst, R. Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011. DOI: 10.1016/j.acha.2010.04.005.

[27] Y. Xu, Q. Q. Kong, W. W. Wang, M. D. Plumbley. Large-scale weakly supervised audio classification using gated convolutional neural network. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, Calgary, Canada, pp. 121−125. 2018. DOI: 10.1109/ICASSP.2018.8461975.

[28] J. B. Zhang, Y. Zheng, J. K. Sun, D. K. Qi. Flow prediction in spatio-temporal networks based on multitask deep learning. *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 3, pp. 468–478, 2020. DOI: 10.1109/TKDE.2019.2891537.

[29] R. Cipolla, A. Kendall, Y. Gal. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Salt Lake City, USA, pp. 7482−7491, 2018. DOI: 10.1109/CVPR.2018.00781.

[30] H. C. Frey, A. Unal, J. Chen, S. Li, C. Xuan. Methodology for Developing Modal Emission Rates for EPA′s Multi-Scale Motor Vehicle & Equipment Emission System, US Environmental Protection Agency, Washington, USA, 2002.

[31] H. C. Frey, B. Liu. Development and Evaluation of Simplified Version of MOVES for Coupling with Traffic Simulation Model, Technical Report, Washington, USA, 2013.

[32] E. J. Keogh, M. J. Pazzani. Derivative dynamic time warping. In *Proceedings of SIAM International Conference on Data Mining*, SIAM, SanJose, USA, pp. 1−11. 2001. DOI: 10.1137/1.9781611972719.1.

[33] J. Liu, W. Guan. A summary of traffic flow forecasting methods. *Journal of Highway and Transportation Research and Development*, vol. 21, no. 3, pp. 82–85, 2004. DOI: 10.3969/j.issn.1002-0268.2004.03.022. (in Chinese)

[34] J. S. Goldstein, J. C. Pevehouse. Reciprocity, bullying, and international cooperation: Time-series analysis of the Bosnia conflict. *American Political Science Review*, vol. 91, no. 3, pp. 515–529, 1997. DOI: 10.2307/2952072.

[35] D. Svozil, V. Kvasnicka, J. Pospichal. Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems*, vol. 39, no. 1, pp. 43–62, 1997. DOI: 10.1016/S0169-7439(97)00061-0.

[36] S. Hochreiter, J. Schmidhuber. Long short-term memory. *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. DOI: 10.1162/neco.1997.9.8.1735.

[37] K. Cho, B. Van Merriënboer, D. Bahdanau, Y. Bengio. On the properties of neural machine translation: Encoder-decoder approaches. In *Proceedings of SSST-8, the 8th Workshop on Syntax, Semantics and Structure in Statistical Translation*, Association for Computational Linguistics, Doha, Qatar, pp. 103−111, 2014. DOI: 10.3115/v1/W14-4012.

Zhen-Yi Zhao received the B. Sc. degree in automation from University of Science and Technology of China, China in 2017. She is currently a Ph. D. degree candidate of control science and engineering in Department of Automation, University of Science and Technology of China, China.

Her research interests include deep learning, urban computing, intelligent transportation, machine learning and data mining.

E-mail: zzy0025@mail.ustc.edu.cn
ORCID iD: 0000-0002-4203-4896

Yang Cao received the B. Sc. and the Ph. D. degrees in information engineering from Northeastern University, China in 1999 and 2004, respectively. Since 2004, he has been with Department of Automation, University of Science and Technology of China, where he is currently an associate professor. He is a member of the IEEE Signal Processing Society.

His research interests include machine learning and computer vision.

E-mail: forrest@ustc.edu.cn
ORCID iD: 0000-0002-2891-4379

Yu Kang received the Ph. D. degree in control theory and control engineering from University of Science and Technology of China, China in 2005. From 2005 to 2007, he was a post-doctoral fellow with Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China. He is currently a professor with Department of Automation, University of Science and Technology of China, China.

His research interests include adaptive/robust control, variable structure control, mobile manipulator, and Markovian jump systems.

E-mail: kangduyu@ustc.edu.cn (Corresponding author)
ORCID iD: 0000-0002-8706-3252

Zhen-Yi Xu received the B. Sc. degree in automation from the Nanjing Institute of Technology, China in 2015, and the Ph. D. degree of control science and engineering from Department of Automation, University of Science and Technology of China, China in 2020. He is currently a post-doctoral fellow with School of Information Science, University of Science and Technology of China, China.

His research interests include deep learning, urban computing, intelligent transportation, machine learning and data mining.

E-mail: xuzhenyi@mail.ustc.edu.cn
ORCID iD: 0000-0002-5804-882X