



OPEN

An urban crowd flow model integrating geographic characteristics

Yu Zhang^{1,2}, Sheng Wu^{1,3,4,6}, Zhiyuan Zhao^{1,3,4,6}✉, Xiping Yang^{5,7} & Zhixiang Fang²

Predicting urban crowd flow spatial distributions plays a critical role in optimizing urban public safety and traffic congestion management. The spatial dependency between regions and the temporal dynamics of the local crowd flow are two important features in urban crowd flow prediction. However, few studies considered geographic characteristic in terms of spatial features. To fill this gap, we propose an urban crowd flow prediction model integrating geographic characteristics (FPM-geo). First, three geographic characteristics, proximity, functional similarity, and road network connectivity, are fused by a residual multigraph convolution network to model the spatial dependency relationship. Then, a long short-term memory network is applied as a framework to integrate both the temporal dynamic patterns of local crowd flow and the spatial dependency between regions. A 4-day mobile phone dataset validates the effectiveness of the proposed method by comparing it with several widely used approaches. The result shows that the root mean square error decreases by 15.37% compared with those of the typical models with the prediction interval at the 15-min level. The prediction error increases with the crowd flow size in a local area. Moreover, the error reaches the top of the morning peak and the evening peak and slopes down to the bottom at night.

Urban crowd flow dynamics are the fundamental data used in many smart city applications, such as urban traffic congestion management, public facility planning, epidemic prevention and control^{1,2}. However, due to the complex nature of human dynamics caused by various spatial and temporal constraints and approaches to overcome these constraints, predicting urban crowd flow is challenging. The current crowd flow in a certain area is jointly affected by the past local crowd flow and the crowd flow in the surrounding areas. Both the temporal dynamic patterns of the local area and the spatial dependency relationships between the local area and the related regions need to be considered for a well-designed urban crowd flow prediction model. Therefore, existing prediction models can be divided into two types: temporal feature-based models and spatial feature-based models.

In terms of the temporal feature-based model, urban crowd flow distribution prediction is modeled as a temporal forecasting problem, and classic time series models are first applied³. The development of machine learning, such as long short-term memory (LSTM), has further improved the accuracy and ability to predict crowd flow based on time relationships⁴. In terms of the spatial feature-based model, modeling and integrating the complex and dynamic spatial relationships between the local areas and the related areas challenges traditional methods. The emergence of deep learning models, such as convolutional neural networks (CNNs)⁵, graph convolutional networks (GCNs)⁶ and residual multigraph convolutional networks (RMGCNs)⁷, has promoted spatial relation modeling and integration. However, in these studies, only some simple relations between adjacent regions (e.g., connectivity relationships) were considered, and many meaningful but complex spatial features have not been well-integrated.

Geological characteristics, such as regional function characteristics and connectivity strength, play important roles in predicting crowd flow. Similar regional functions between areas imply similar crowd flow dynamics rhythm. For example, the crowd flow in a residential area exhibits substantial loss in the morning and recovers at night. As a result, local crowd flow dynamic patterns can contribute to improving the prediction performance for crowd flow in areas with similar regional functions. Moreover, connectivity strength can effectively measure the

¹Academy of Digital China (Fujian), Fuzhou University, Fuzhou, China. ²State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China. ³Key Laboratory of Spatial Data Mining & Information Sharing of Ministry of Education, Fuzhou, China. ⁴Ministry of Education Fujian Collaborative Innovation Center for Big Data Applications in Governments, Fuzhou, China. ⁵School of Geography and Tourism, Shaanxi Normal University, Xi'an, China. ⁶The Digital Economy Alliance of Fujian, Fuzhou, China. ⁷Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources, Shenzhen, China. ✉email: zyzhao@fzu.edu.cn

impacts from both adjacent regions and distant areas. For example, urban expressways can considerably improve the connectivity strength between distant regions by reducing the travel time. Therefore, only considering the adjacent areas to model the spatial relationships may ignore the impacts of crowd flow from distant but highly connected areas. A method that can integrate multiple spatiotemporal relationships is needed to improve the urban crowd flow prediction performance.

In this study, we aim to answer the following two questions: (1) How do geographic characteristics affect the prediction of urban crowd flow? (2) What is the performance of the proposed method compared to the method without considering geographic characteristics (e.g., LSTM, RMGCN) and how are the prediction errors distributed on the spatial and temporal dimensions?

This paper is organized as follows: In the related work section, a comprehensive overview is given. Then, the methodology of this paper is introduced. Next, the experiments and results, the experimental data and result analysis are presented. We discuss related results and draw several conclusions in the final section.

Related work

According to the parametric requirements, existing urban crowd flow prediction models can be divided into two types: parametric and nonparametric. The parametric model is built based on a regression function, and the unknown parameters need to be estimated based on the benchmark dataset before predicting the regional crowd flow. For example, Pappalardo et al. proposed an analytical framework to nowcast the population count, well-being, and economic development based on mobile phone data⁸. Classic models, such as the historical average model⁹, autoregressive integrated moving average model (ARIMA)¹⁰ and Kalman filter¹¹ all belong to this type. The historical average model uses the average value of traffic over historical moments as the predicted value. The ARIMA model proposed by Box and Jenkins¹² regards the crowd flow dynamics in a local area as a time series and predicts the crowd flow by curve fitting and parameter estimation based on historical data. The Kalman filter establishes a linear system state equation and uses the crowd flow at the previous moment to obtain a prediction. However, traditional parametric models are not adept at addressing features of complexity and uncertainty (e.g., nonlinear patterns between variables) in urban crowd flow prediction problems.

The nonparametric model requires few basic assumptions between available variables and outperforms traditional parametric models in addressing complex features. The relationships between the dependent variables and the independent variables are learned and derived from historical data to establish an approximate model¹³. Typical approaches include K-nearest neighbors (KNN)¹⁴, Bayes¹⁵, support vector regression (SVR)¹⁶ and deep learning^{17,18}, which have been widely used in urban crowd flow prediction applications as well as many other related fields. Recently, the deep learning model has developed rapidly due to its outstanding performance in integrating temporal and spatial characteristics of crowd flow dynamics¹⁹.

Integrating different spatial and temporal features with different matched deep learning models results in a varying performance. Early neural network models only considered temporal characteristics^{20,21}. For example, Fu et al. used an LSTM model and a gated recurrent unit model to predict traffic flow²². Tian et al. also used an LSTM model to predict traffic speed and traffic flow. To better integrate spatial features, many deep learning models have been widely investigated and applied^{23–25}. Chen et al. used an artificial neural network to predict the population in each grid from a neighborhood perspective²⁶. Guangyuan et al. used a convolution LSTM model to predict the spatiotemporal distribution of mobile phone users at a fine-grained temporal resolution²⁷. Zhang et al. proposed a spatial and temporal residential network model to predict crowd flows⁷. Wu et al. combined a CNN and LSTM to model the spatial and temporal relationships between regions, respectively, to construct a prediction model²⁸. In terms of the spatial relationships, GCNs exhibit better modeling performance²⁹. For example, Chai et al. used a GCN model to construct various spatial relationships between stations to predict bicycle traffic³⁰. Sun et al. proposed a multi-view GCN (MVGCN) model to construct spatial correlations and interactions between irregular regions to predict crowd flow³¹.

However, many meaningful geographical features are rarely considered in spatial features during the above studies. The prediction errors and the effectiveness of different features lack insightful investigation and comprehension.

Methodology

Problem definition. We divide the study area into regular grids of equal size, and each grid is represented by v_i ; $V = \{v_1, v_2, \dots, v_N\}$ is the set of all the grids, and N represents the total number of grids, where $i \in [1, N]$. A graph $G = (V, A)$ is constructed to represent the spatial relationships between grids. Each grid is regarded as a node, and the spatial link between each pair of grids is regarded as an edge. $A \in \mathbf{R}^{N \times N}$ represents the spatial relation matrix between grids. $X_t \in \mathbf{R}^N$ represents the crowd flow for all grids at time t . Predicting urban crowd flow can be modeled by constructing the function $F(\cdot)$ to calculate X_t based on the spatial relation matrix A and the historical urban crowd flow $X_{t-1}, X_{t-2}, X_{t-H+1}$ in a time window H (Eq. 1).

$$X_t = F(A; (X_{t-1}, X_{t-2}, X_{t-H+1})) \quad (1)$$

The framework of the proposed method. As shown in Supplementary Fig. 1, we first use mobile phone data to extract the crowd flow in each grid at every moment. Then, we propose three geographic characteristics to measure the spatial dependence from different perspectives and adopt the RMGCN to model these three features. Finally, an LSTM model is applied to integrate dependence relationships on both the temporal and spatial dimensions.

Graph construction. Three graphs were constructed based on geographic characteristics from different perspectives (Supplementary Fig. 2): proximity, functional similarity, and connectivity relationships.

The proximity is measured by the adjacent relation. The proximity characteristic is derived from Tobler’s first law (TFL), which states that “everything is related to everything else, but near things are more related than distant things”³². We believe that the crowd flow in a grid could be strongly affected by nearby grids. For intuitive purposes, we adopt the Moore neighborhood (i.e., the eight grids that surround the central grid) to measure the proximity relationship (Eq. 2).

$$A_{ij}^J = \begin{cases} 1, & \text{if } v^j \text{ belongs to the 8 grids that surround } v^i \\ 0, & \text{else} \end{cases} \tag{2}$$

where A_{ij}^J represents the proximity between v^i and v^j .

The functional similarity is measured based on the similarity of the geographical properties. The functional similarity characteristic is derived from the “Third Law of Geography”, which states that “the more similar geographic configurations of two points (areas), the more similar the values (processes) of the target variable at these two points (areas)”³³. According to this law, areas with similar context will have similar features. As a result, the crowd flow tends to be similar for grids with similar urban functions. For example, the crowd flow in areas with middle schools or office buildings share similar crowd flow patterns. Considering the accessibility of the data, we adopt the shared structure of different point of interest (POI) categories to represent the local geographical property. A simple method is constructed based on the POI shared structure to measure the functional similarity. In terms of the calculation method, we regard each grid as a vertex and measure the similarities of the functions between the grids (Eq. 3).

$$A_{ij}^P = sim(\mathit{poi}^i, \mathit{poi}^j) \tag{3}$$

The connectivity is measured by the shortest path distances between regions based on the road network. The connectivity characteristic is derived from the combination of the “Second Law of Geography” and the TFL. The “Second Law of Geography” implies the spatial heterogeneity of geographical phenomena and states that “geographic variables exhibit uncontrolled variance”³⁴. According to this law, the effectiveness of a spatial model integrating geographical features relies on the location of the analysis areas, especially distant areas. However, the connectivity characteristics (e.g., express road in two cities) mean that distant areas can affect each other directly, making the TFL valid again. A shorter distance indicates a stronger connectivity relation. To improve the calculation efficiency, we regard two grids as unconnected if their shortest path distance exceeds threshold δ . Threshold δ can be set by referring to the prediction time step. In addition, considering that the grids surrounding a local grid have been modeled by the proximity relationship, these grids are not considered. The calculation equation is as follows:

$$A_{i,j}^C = \begin{cases} \frac{1}{d(v^i, v^j)}, & 0 < d(v_i, v_j) < \delta \quad \text{and} \quad A_{i,j}^J = 0 \\ 0, & \text{esle} \end{cases} \tag{4}$$

$$d(v^i, v^j) = \min(\mathit{dist}(v^i, v^j)) \tag{5}$$

where $\mathit{dist}(v^i, v^j)$ represents the Euclidean distance between v^i and v^j , and $d(v^i, v^j)$ is the shortest distance, \min is the minimum function, and $A_{i,j}^C$ represents the connectivity strength.

Graph fusion. To better model the geographic characteristics between grids, we need to perform graph fusion and merge graphs composed of different types of spatial relations into one graph. We perform graph fusion using weighted summation. We first normalize the three spatial relation matrices to handle the large value differences between different graphs (Eq. 6).

$$L = I - D^{1/2}AD^{1/2} \tag{6}$$

$$DD = \begin{bmatrix} \sum_{j=0}^{n-1} A_{0,j} & 0 & \cdots & 0 \\ 0 & \sum_{j=0}^{n-1} A_{1,j} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sum_{j=0}^{n-1} A_{n-1,j} \end{bmatrix} \tag{7}$$

where A represents the spatial relation matrix ($A \in [A^J, A^P, A^C]$), D represents the corresponding degree matrix, and the calculation method is shown in Eq. (7). L represents the normalized spatial relation matrix. I is the identity matrix.

We normalize the fused spatial relation matrices according to Eq. (8) to obtain the normalized proximity matrix L^J , functional similarity matrix L^P , and road network connectivity matrix L^C . Then, the weighted summation of these three normalized spatial relation matrices is calculated.

$$L^{Fu} = W^0 \odot L^J + W^1 \odot L^P + W^2 \odot L^C \tag{8}$$

where L^{Fu} is the spatial relation matrix after weighted summation. W^0 , W^1 , and W^2 are trainable parameters.

Residual graph convolution. To better capture the geographic characteristics between regions, we use a residual GCN for modeling purposes. As shown in Supplementary Fig. 3, the residual graph convolution model is constructed based on the graph convolution model, but a layer of residual links is added to each layer. These residual links add the input and output of each layer as the next layer (Eq. 12) because the transmission of information and gradients provides an additional connection channel, which can resolve the smoothness problem. The chosen graph convolution model is a graph convolution model of order k based on Chebyshev polynomials³⁵. K determines the scope of the graph convolution. As shown in Supplementary Fig. 4, when $K=0$, $T_0(L)=I$, and this model represents only the node itself. When $K>0$, the model can extract the first-order to K th-order neighborhood relations of the predicted grid. For example, when $K=2$, the model can extract the first-order and second-order neighborhood relations of the predicted grid as follows:

$$X_{l+1} = \text{relu} \left(\sum_{k=0}^K \alpha_k T_k(L^{Fu}) X_l \right) \tag{9}$$

$$\begin{cases} T_0(L) = I \\ T_1(L) = L^{Fu} \\ T_2(L) = 2T_1(L^{Fu}) - T_0(L^{Fu}) \\ \dots \\ T_k(L) = 2T_{k-1}(L^{Fu}) - T_{k-2}(L^{Fu}) \end{cases} \tag{10}$$

$$\text{relu}(x) = \max(0, x) \tag{11}$$

$$X'_{l+1} = X_{l+1} + X_l \tag{12}$$

where X_l is the input feature of layer l of N grids. X_{l+1} is the output feature of layer l of N grids. X'_{l+1} is the input feature of the $l+1$ layer of the N grids. M is the number of features. $T_k(\cdot)$ is a Chebyshev polynomial (Eq. 10). α_k is a trainable parameter. relu is the rectified linear unit activation function (Eq. 11), and \max is the maximum function.

Temporal dependence modeling. To determine the influence of historical crowd flow dynamics, we use the LSTM model to extract the temporal characteristics³⁶. Instead of using the historical crowd flow as input directly for the traditional LSTM model, we use the result of the residential GCNs in the previous step as the input. An LSTM unit consists of a cell and three gates: an input gate, an output gate, and a forget gate. The cell state saves the crowd flow information in this study. The specific calculation process is as follows:

First, the crowd flow matrix X_t of each grid at time t is input as the origin state of the cell, and the output h_{t-1} at time $t-1$ and the input X_t at the current time t are used to calculate the forget gate f_t . The output of this layer is a value between 0 and 1, which is used to determine the crowd flow information retention degree at time $t-1$.

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f) \tag{13}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{14}$$

where h_{t-1} represents the output at time $t-1$, which is obtained through an iterative loop calculation. For details, please refer to Eq. (20) in the last step of this process. f_t indicates the forget gate function at time t . W_f is the weight matrix of the input layer, and the optimal value is obtained through model training. b_f is the paranoid item of the input layer, and the optimal value is also obtained through model training. σ is the sigmoid function (Eq. 14).

Then, the output h_{t-1} at time $t-1$ and the input X_t at time t are used to calculate the input gate i_t . The output of this layer is a value between 0 and 1, which is used to determine the degree of retention of the crowd flow at time t . In addition, the output h_{t-1} at time $t-1$ and the input X_t at the current time t are used to generate a candidate vector \tilde{C}_t as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \tag{15}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, X_t] + b_c) \tag{16}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{17}$$

where W_i and W_C represent the weight matrices in the input and state update layers, respectively, b_i and b_c are corresponding paranoid terms, and the optimal values are obtained through model training. \tanh is the activation function (Eq. 17).

Then, the cell state is updated: C_{t-1} is updated to C_t . The value f_t of the forget gate is multiplied by the old cell state C_{t-1} , which stores historical crowd flow information, and part of the historical crowd flow information is forgotten. Then, the input gate value i_t is multiplied by the candidate vector \tilde{C}_t to store part of the crowd flow information at the current moment. Finally, the two results are added together to determine the new cell state as follows:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (18)$$

Last, the output gate o_t is calculated using the output h_{t-1} at time $t-1$ and the input X_t at time t . The output of this layer is a value between 0 and 1, which is used to determine which parts of the new cell state will be output. Then, the \tanh function is used to process the cell state C_t , and the processed value is multiplied by the output gate value o_t to obtain the output value as follows:

$$o_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \quad (19)$$

$$h_t = o_t * \tanh(C_t) \quad (20)$$

where W_o and b_o are the weight matrix and the paranoid term of the input layer, respectively, and the optimal value is obtained through model training.

Evaluation and validation. We use the mean absolute error (MAE) and root mean square error (RMSE) and mean absolute percentage error (MAPE) to evaluate the outcomes of the proposed model, as well as other comparative models. MAE refers to the mean value of the absolute error between the predicted value and the actual observed value (Eq. 21). RMSE is the square root of the mean square difference between the predicted value and the actual observed value (Eq. 22). MAPE is a measure of the percentage error of the forecast in relation to the actual observed values (Eq. 23).

$$MAE^i = \frac{\sum_{t=1}^Q |y_t^i - y_t^{i'}|}{Q} \quad (21)$$

$$RMSE^i = \sqrt{\frac{\sum_{t=1}^Q (y_t^i - y_t^{i'})^2}{Q}} \quad (22)$$

$$MAPE^i = \frac{\sum_{t=1}^Q \frac{|y_t^i - y_t^{i'}|}{y_t^i}}{Q} \quad (23)$$

where y_t^i and $y_t^{i'}$ represent the observed and predicted crowd flows of grid i at time t , respectively. Q is the predicted sample.

To better deal with the skewness problem (e.g., the Simpson paradox³⁷), we added the weighted mean absolute error (WMAE) and the weighted mean absolute percentage error (WMAPE) to evaluate the results.

$$WMAE = \sum_{i=1}^N |y_t^{i'} - y_t^i| \quad (24)$$

$$WMAPE = \frac{WMAE}{\sum_{i=1}^N |y_t^i|} \quad (25)$$

Experiments and results

Data and data preprocessing. A mobile phone location dataset of Xining, a city in western China, is adopted in this study. The dataset contains more than 170 million records collected from approximately 0.3 million mobile phone users. This dataset is generated by the incoming or initiated actions of the user and contains a variety of record types (e.g., calls, SMS, internet access, etc.). Each record includes a corresponding timestamp, location and anonymized user ID. The users account for 22% of the population of the city. The dataset covers 4 consecutive work days in August 2018. Each user has 140 records a day on average, which indicates an average temporal interval shorter than 15 min. We extract 2559 base stations from the dataset (Supplementary Fig. 5), of which approximately 96% cover less than 500 m (Supplementary Fig. 6).

The POI data are applied to describe the region function and measure the function similarity between regions, which is A^P in Section "Graph construction". The data are collected based on the application programming interface (API) of Amap (www.amap.com), which is one of the most popular online map service products in China (Supplementary Fig. 7(a)). The dataset contains 21 categories, such as companies and scenic spots. (Supplementary Table 1).

A road network is applied to measure the connectivity strength between every two regions and construct a spatial connectivity matrix A^c in Section "Graph construction". The road network data (Supplementary Fig. 7(b)) are downloaded from the OpenStreetMap (www.openstreetmap.com). To reduce the impacts from unevenly distributed small roads, only the expressway, the main road and the secondary road are retained.

To build and examine the proposed model, we first calculate the crowd flow of each grid. To avoid the potential impacts caused by the substantial crowd flow difference between the day and the night derived from the origin dataset, we need to estimate the location of each user at the target timestamps.

First, we divide a day into several fixed time windows and estimate the most likely location of each user in every time window. The mobile phone location records are distributed unevenly in the temporal dimension. In particular, a time window for a user trajectory may contain no record or may have more than one record. To better estimate the most likely location, the record closest to the middle timestamp is retained by referring to the method in Zhao et al.³⁸.

Second, we divide the study area into grids and calculate the crowd flow of each grid. The grids are defined based on the left-lower corner location and the fixed size. Each grid is coded by corresponding row and column numbers. For every time window, we calculate the crowd flow by counting the user number for each grid based on its colocation relation.

For the POIs in each grid, the proportions of the 21 categories are calculated. We use the Kolmogorov–Smirnov test (K-S test) to investigate the fitness of the normal distribution and the POI distribution for the grid. The p-value is 0.037, which indicates a normal distribution at the 0.1 significance level but not at the 0.05 or smaller significance level. We further test the distribution by category for each grid and find that 79% of the grid exhibits a normal distribution at the 0.05 significance level. The above results imply that the POI distribution satisfies a normal distribution, and the Pearson correlation coefficient can reasonably describe the similarity of the POIs between grids in general. Therefore, we set the similarity algorithm in Eq. (3) to the Pearson correlation coefficient.

$$\text{sim}(\mathbf{poi}^i, \mathbf{poi}^j) = \begin{cases} r_{ij}, & r_{ij} > 0 \\ 0, & r_{ij} \leq 0 \end{cases} \quad (26)$$

$$r_{ij} = \frac{\sum_{z=1}^Z (\text{poi}_z^i - \overline{\text{poi}}^i)(\text{poi}_z^j - \overline{\text{poi}}^j)}{\sqrt{\sum_{z=1}^Z (\text{poi}_z^i - \overline{\text{poi}}^i)^2} \sqrt{\sum_{z=1}^Z (\text{poi}_z^j - \overline{\text{poi}}^j)^2}} \in [-1, 1] \quad (27)$$

where poi_z^i and poi_z^j represent the ratio of POI type z in grid i and grid j , respectively; $z \in [1, Z]$; Z represents the type number of all POIs. A_{ij}^p represents the functional similarity between v^j and v^i . As we said before, grids with the same function will improve the accuracy of the prediction grid.

Referring to existing research³⁹, we set the negative r value as zero.

We use the network analyst module of ArcGIS 10.2 to organize the road network and calculate the shortest path between the grids. Then, the connectivity strength matrix can be built by Eqs. (4) and (5).

Parameter settings. In the experiment, we choose the data from the first three days as the training set and the data from the last day as the test set. The grid size, the time window and the connectivity threshold are three basic parameters that need to be set. First, we choose 500 m to define the grid size by jointly considering the spatial resolution of the dataset and the commonly used size in previous studies^{40,41}. Second, considering the temporal intervals of the mobile phone dataset, we set the most fine time window as 15 min and further analyzed and compared the outcomes for 30 min, 45 min, and 60 min. Third, the distance threshold δ in the connectivity graph is set based on the daily movement speed (i.e., 80 km/h in this study) in urban areas and the predicted time window width.

The parameters of the proposed FPM-geo model mainly include the learning rate, batch processing volume, number of training iterations, number of layers and graph convolution order. In this experiment, the learning rate is set to 0.001, the batch size is set to 16, the number of training iterations is set to 2000, the number of graph convolution layers is set to 3, and the graph convolution order is set to 2 (Supplementary Fig. 8).

Prediction results. Figure 1 shows the results of the urban crowd flow prediction with a 500-m grid size and a 15-min prediction step. The urban crowd flow is concentrated mainly along the two cross rivers in the downtown area. The absolute prediction differences during the night tend to be smaller than those during the daytime. Moreover, the grids with large absolute prediction differences are distributed more sparsely during the morning and evening rush hours and concentrate on the center area during the working hours.

Comparative analysis. To better reflect the prediction performance, 4 prediction time steps are applied, namely, 15 min, 30 min, 45 min, and 60 min. We compare the proposed model (FPM-geo) with the following 5 prediction methods: SVR is a multivariate extension of the autoregressive model, which is capable of modeling correlations between regions. LSTM is a variant of traditional RNN. It can effectively capture the semantic association between long sequences and alleviate the phenomenon of gradient disappearance or explosion compared with classical RNN. In GCN, features are extracted from graph data and used to make predictions. TGCN and STMGCN are both graph convolution based traffic prediction models. TGCN only considers a single spatial relationship. STMGCN considers multiple spatial relationships.

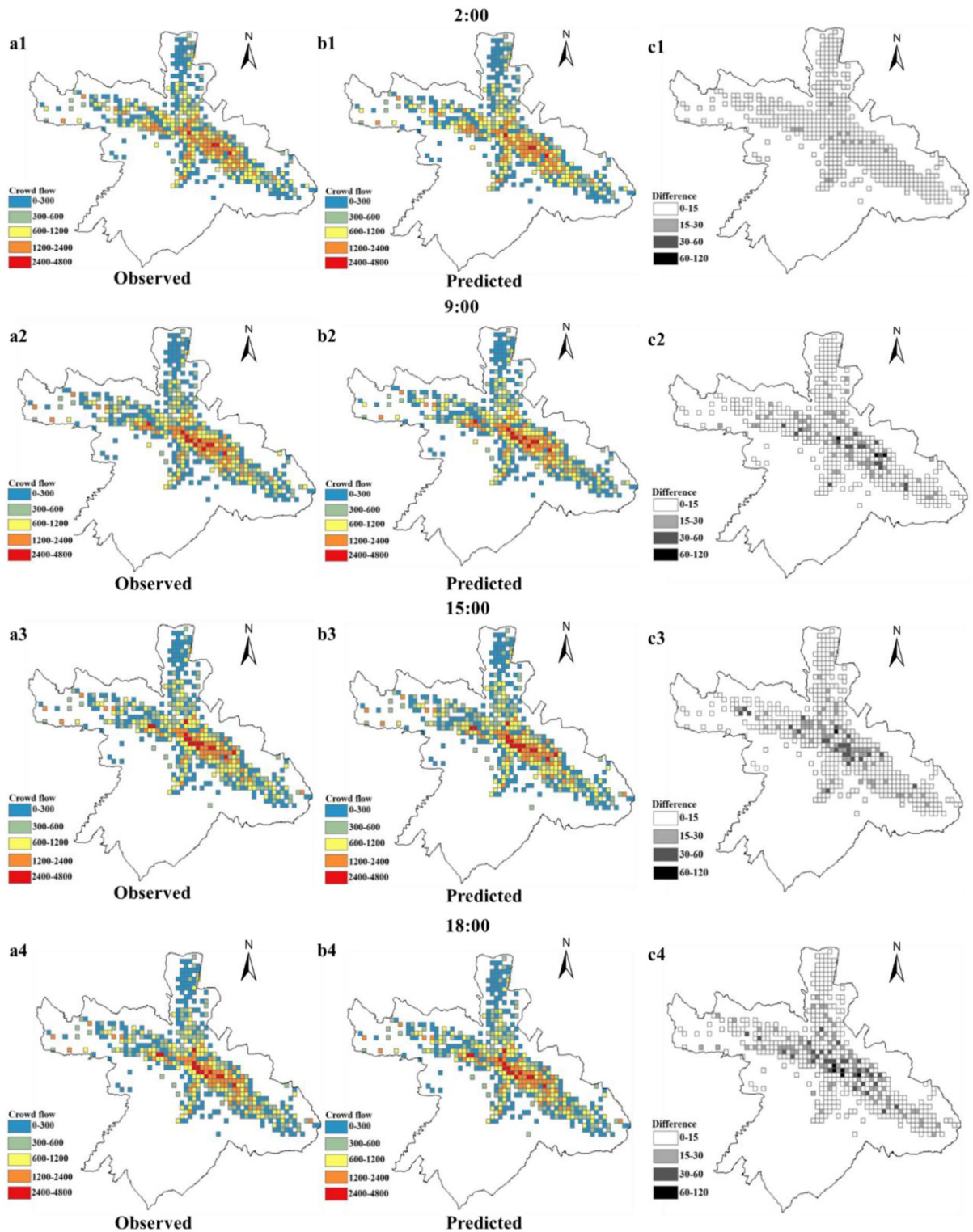


Figure 1. Crowd flow prediction results.

The result indicates that the proposed FPM-geo outperforms all 5 selected models for each prediction time step (Table 1). Among these models, machine learning models (i.e., SVR) have the lowest prediction accuracy. Compared with the model that only considers temporal features (i.e., LSTM) or the model that only considers spatial features (i.e., GCN), the models that integrate both spatial and temporal features (i.e., TGCN, STMGCN) have higher prediction accuracy. For the model that only considers limited spatial and temporal features (i.e., STMGCN), our model achieves a better prediction performance by integrating multiple geographical features.

T (min)	Metric	SVR	LSTM	GCN	TGCN	STMGCN	FPM-geo
15	MAE	12.56	9.46	8.93	8.72	8.25	7.90
	RMSE	20.86	16.18	14.89	14.70	13.05	12.44
	MAPE	0.0431	0.0338	0.0333	0.0327	0.0326	0.0304
	WMAE	20.87	15.89	15.02	14.52	13.42	12.53
	WMAPE	0.0170	0.0130	0.0122	0.0118	0.0109	0.0101
30	MAE	19.26	14.42	14.06	12.99	12.26	12.18
	RMSE	33.51	25.83	24.62	22.82	20.81	19.93
	MAPE	0.0645	0.0492	0.0450	0.0439	0.0431	0.0427
	WMAE	42.99	25.67	24.84	23.09	20.56	19.97
	WMAPE	0.0352	0.0210	0.0203	0.0188	0.0167	0.0162
45	MAE	25.67	18.99	18.25	17.26	16.24	15.45
	RMSE	47.44	35.11	32.29	30.53	27.82	26.76
	MAPE	0.0780	0.0545	0.0527	0.0526	0.0515	0.0502
	WMAE	56.58	33.67	31.77	30.69	27.97	25.82
	WMAPE	0.0463	0.0275	0.0259	0.0250	0.0228	0.0210
60	MAE	30.22	23.55	22.92	21.50	20.5	19.31
	RMSE	60.47	43.88	40.30	39.78	36.13	34.95
	MAPE	0.0909	0.0684	0.0635	0.0611	0.0599	0.0588
	WMAE	69.56	42.67	41.21	40.10	37.79	33.15
	WMAPE	0.0569	0.0348	0.0337	0.0327	0.0309	0.0270

Table 1. Performance comparison among different prediction models.

For example, the FPM-geo outperforms the STMGCN and LSTM with decreased RMSE values of 4.67% and 23.12%, respectively, with a prediction time step of 15 min.

The proposed FPM-geo model exhibits stronger robustness in MAE, RMSE, MAPE, WMAE and WMAPE than the other models by the changes in the prediction steps. The MAE, RMSE, MAPE, WMAE and WMAPE values of each model increase with the prediction time window. However, the prediction errors of the FPM-geo model are lower than those of the other models. In particular, when the prediction step lengths are 45 min and 60 min, the MAE, RMSE, MAPE, WMAE and WMAPE values of the FPM-geo model are considerably lower than those of the other models.

The effects of spatial relationships. The three typical geographic characteristics make the main contribution to the effectiveness of the proposed model, especially for areas or time periods with complex population flows. Each geographic characteristic contributes varying effectiveness for different places. To further investigate the effectiveness of the three geographic characteristics, we compare the prediction performance of the pure LSTM (no geographic characteristic is considered), the original FPM-geo (all three geographic characteristics are integrated) and the modified FPM-geo by removing different geographic characteristics. The FPM-geo-J, FPM-geo-F, and FPM-geo-C in Table 2 indicate the modified models in which the proximity relationship (J), the functional similarity relationship (F) and the road connectivity relationship (C) are removed, respectively. Table 2 indicates that the LSTM model prediction errors are the largest, and each modified model has smaller errors than the LSTM but larger errors than the FPM-geo model. The results imply that each spatial relationship contributes a positive effect on the prediction performance.

To further compare the contribution of each geographic characteristic to crowd flow prediction for different places, we select three typical regions, i.e., the city center (Region 1), suburbs (Region 2) and outer suburbs (Region 3), for further analysis (Fig. 2). As shown in Figs. 3, 4 and 5, for the city center, removing the proximity

Method	MAE	RMSE	MAPE
LSTM	9.46	16.18	0.0334
FPM-geo-J	8.37	13.42	0.0326
FPM-geo-F	8.20	12.79	0.0324
FPM-geo-C	8.17	13.01	0.0324
FPM-geo	7.9	12.44	0.0304

Table 2. Influence of spatial relationships on prediction. In this table, FPM-geo-J is the model without proximity. FPM-geo-F is the model without functional similarity. FPM-geo-C is the model without road network connectivity.

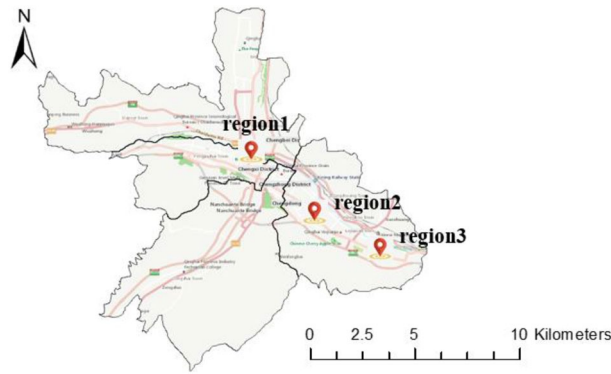


Figure 2. Three typical regions.

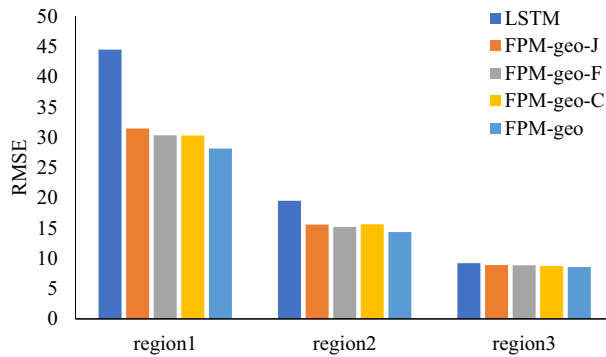


Figure 3. RMSEs of typical regions with different considered spatial relationships.

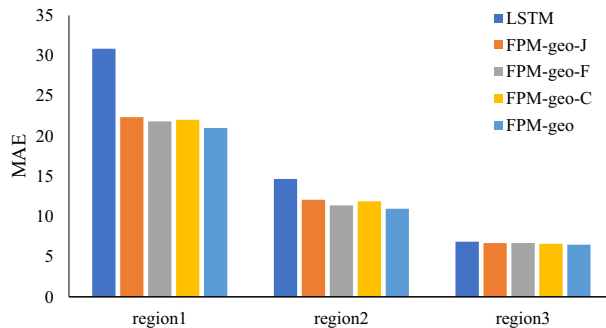


Figure 4. MAEs of typical regions with different considered spatial relationships.

relationship (FPM-geo-J) results in less improvement to the LSTM than that for the other two modified models. This indicates that the proximity relationship plays a more critical role in the crowd flow prediction model in the urban center area (Region 1). Similarly, the connectivity and proximity of the road network make greater contributions in suburban areas (Region 2). For the outer suburbs, the overall crowd flow is small, and there are very few people moving, so the three geographic characteristics have relatively small impacts.

The spatial distribution of the prediction error. The spatial distributions of the prediction errors are shown in Figs. 6, 7 and 8. The errors of the proposed FPM-geo are lower than those of the other two methods for most grids (Figs. 9, 10 and 11). For the downtown areas with high crowd flow density located at the center of the city, the improvement is more obvious. In terms of the spatial distribution, the grids with high RMSE and MAE are mainly concentrated in the central areas and the western part of the city (Figs. 6 and 7). These areas share a common characteristic of high crowd flow density. Residential, office and commercial buildings concentrate in limited space and mix with each other. The crowd flow flows vary in travel purpose (e.g., commuting travel

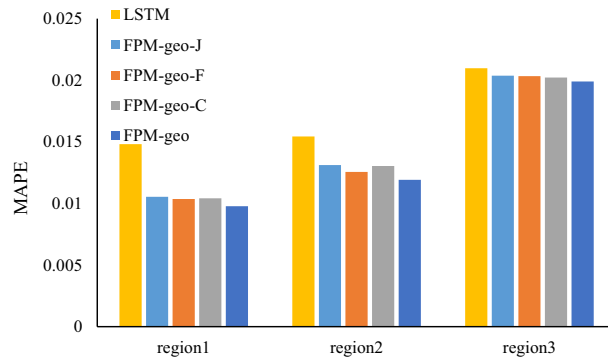


Figure 5. MAPEs of typical regions with different considered spatial relationships.

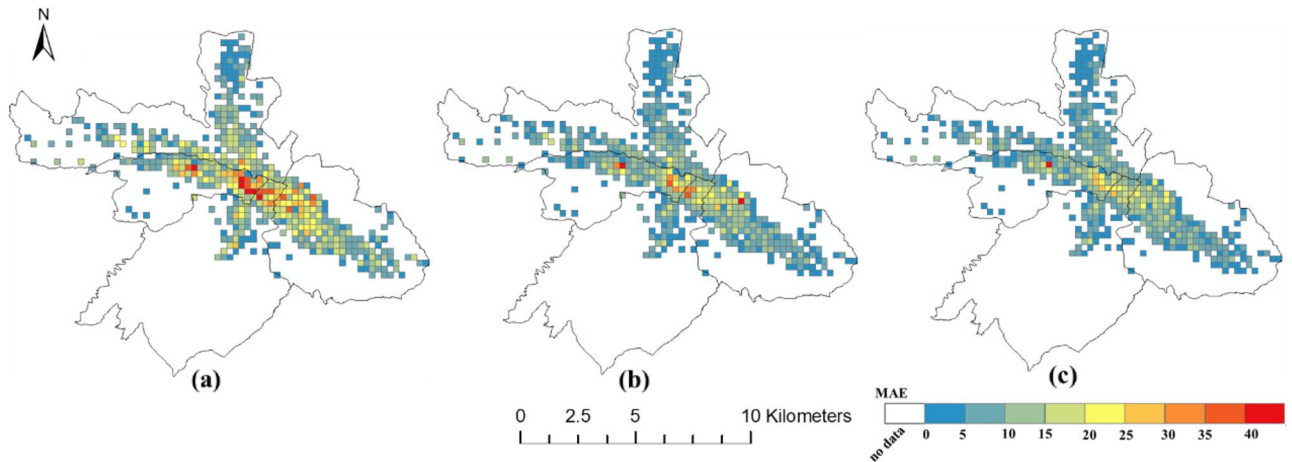


Figure 6. Spatial distributions of the MAEs of different models: (a) SVR; (b) TGCN; (c) FPM-geo.

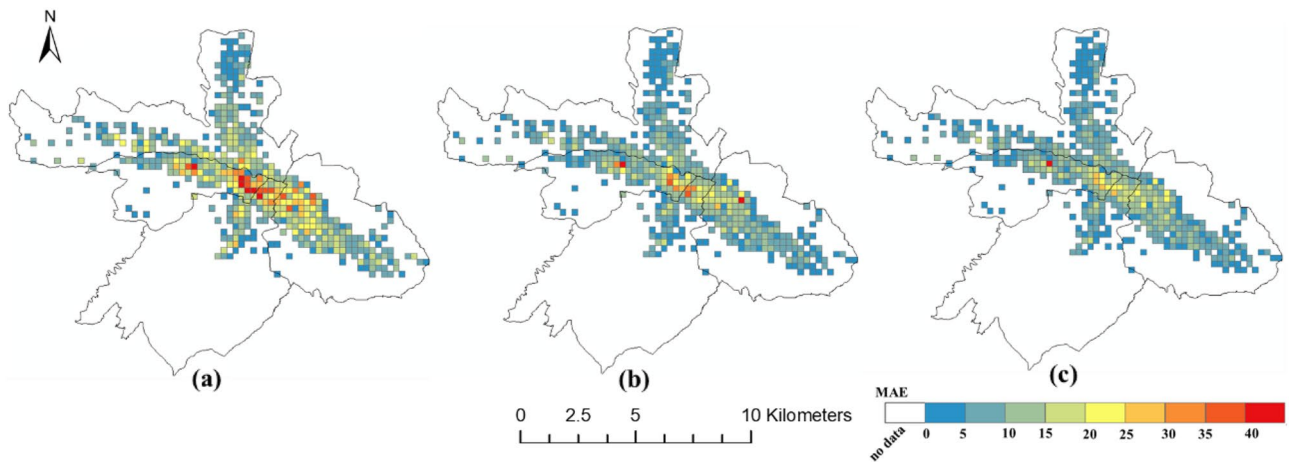


Figure 7. Spatial distributions of the RMSEs of different models: (a) SVR; (b) TGCN; (c) FPM-geo.

and entertainment travel), as well as travel distance (e.g., short travel from local citizens and long travel from suburban citizens). Therefore, predicting the crowd flow in these areas needs to consider the complex nature of the flow. This is the main reason why the proposed FPM-geo outperforms the other methods in this study, especially for downtown areas. For the grids with low crowd flow density located in suburban areas, the crowd flow is relatively simple. Both the proposed method and the comparative methods share a low and similar performance. In contrast to RMSE and MAE, grids with higher MAPE are mostly distributed in suburbs with less crowd flow (Fig. 8). This indicates that high prediction errors mainly occur in grids with small crowd flows. The risk of

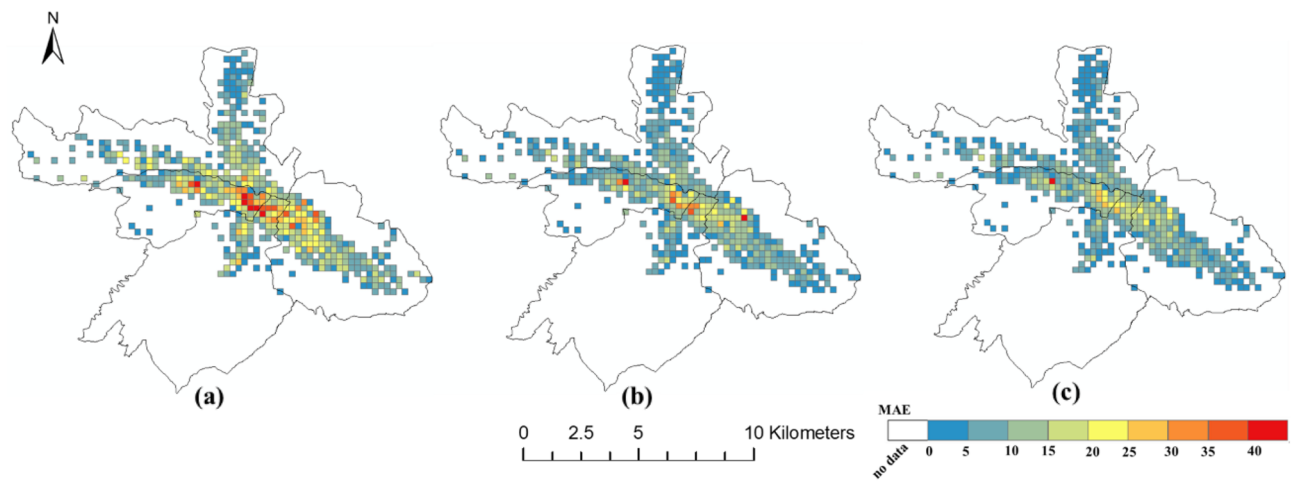


Figure 8. Spatial distributions of the MAPEs of different models: (a) SVR; (b) TGCN; (c) FPM-geo.

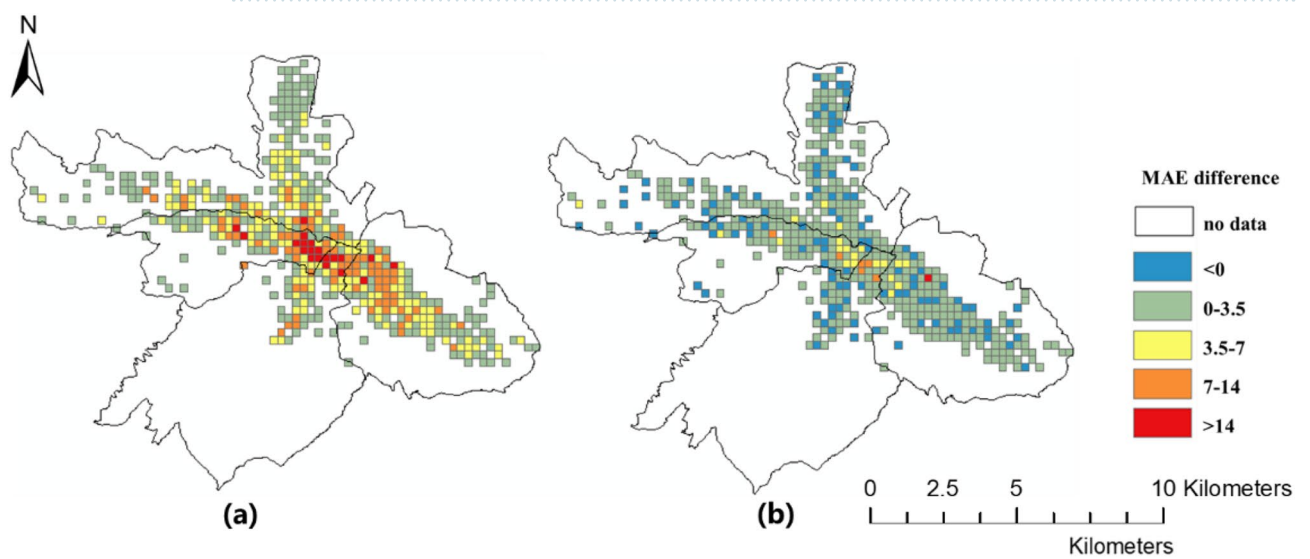


Figure 9. Spatial distributions of the MAE differences among different models: (a) FPM-geo and SVR; (b) FPM-geo and TGCN.

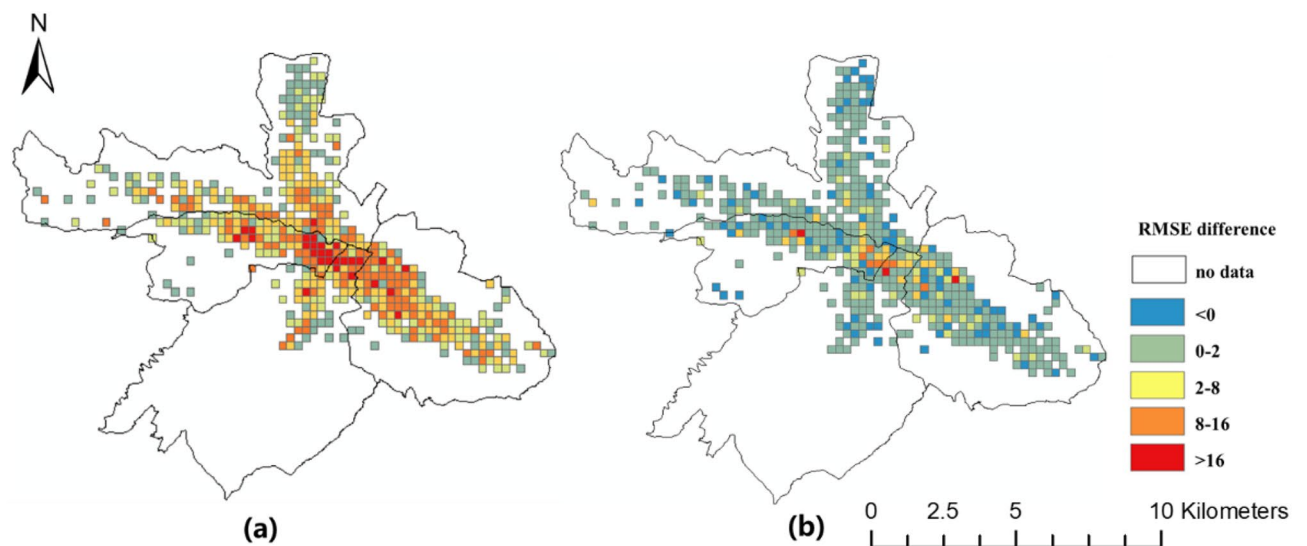


Figure 10. Spatial distributions of the RMSE differences among different models: (a) FPM-geo and SVR; (b) FPM-geo and TGCN.

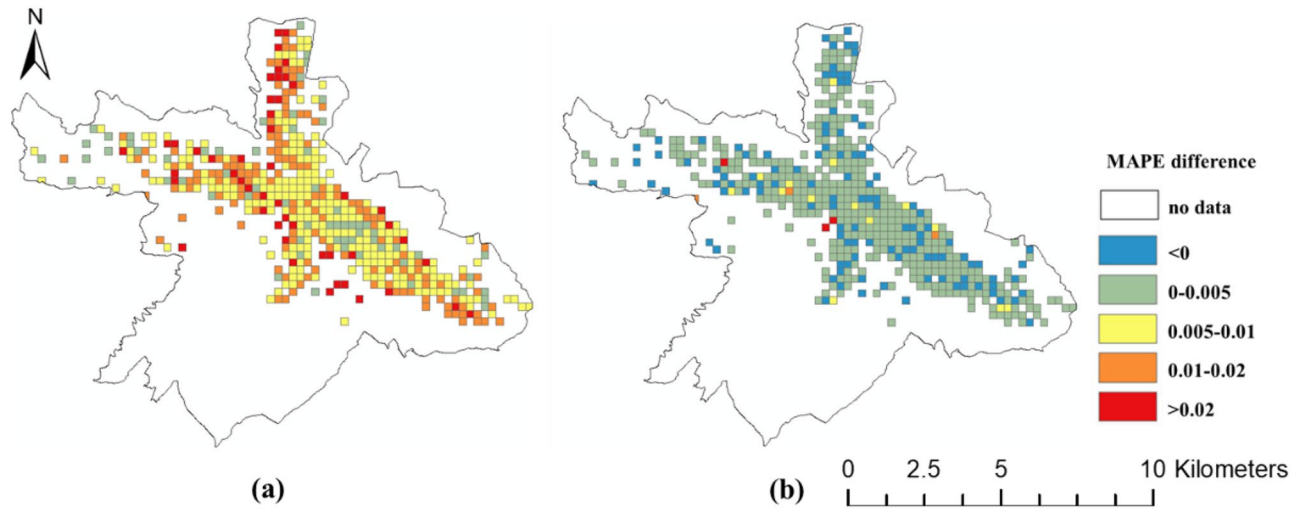


Figure 11. Spatial distributions of the MAPE differences among different models: (a) FPM-geo and SVR; (b) FPM-geo and TGCN.

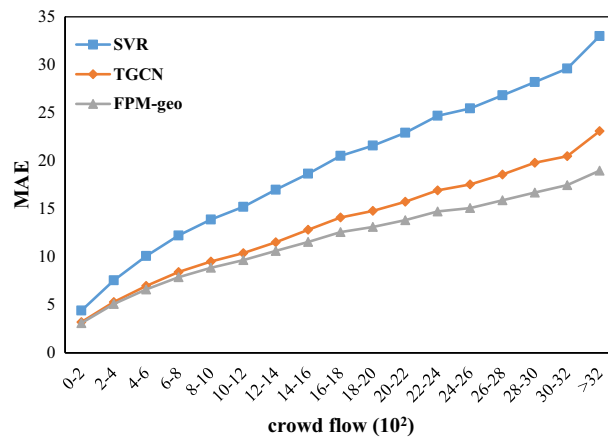


Figure 12. MAE distributions based on the number of people in the grid.

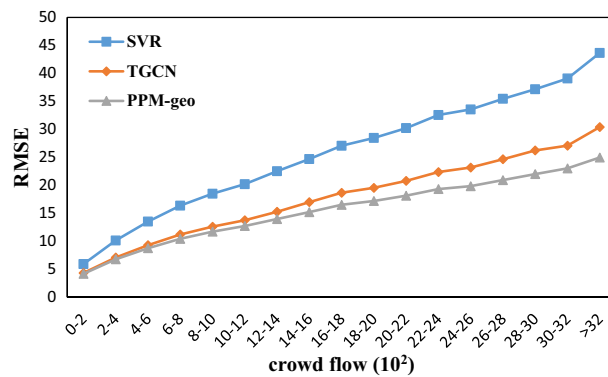


Figure 13. RMSE distributions based on the number of people in the grid.

emergency urban events is more prevalent in densely populated areas. The MAPEs of several models are smaller on grids with high crowd flow, indicating the effectiveness of several models in practical applications. In areas with large crowd flows, FPM-geo also outperforms the other methods, which further reflects its performance.

The spatial distribution of the MAE and RMSE implies a positive relation between the prediction errors and the crowd flow. Further investigation shows that both the MAE and the RMSE of the three models increase with

the crowd flow (Figs. 12 and 13). However, the FPM-geo outperforms the other two methods in both absolute prediction errors and the robustness of the good performance with the increase in crowd flow. For example, when the crowd flow changes from the range [0, 200] to the range [2000, 2200], the RMSE of FPM-geo increases from 4.17 to 18.13, TGCN increases from 4.32 to 20.78, and SVR increases from 5.91 to 30.21. The nearly 2000 increase in crowd flow leads to 13.96, 16.46 and 24.70 decreases for the FPM-geo, TGCN and SVR models, respectively. The FPM-geo exhibits 84.81% and 56.52% of the prediction performance loss caused by the same crowd flow increase for the TGCN and SVR models, respectively. When the crowd flow reaches 2000, the RMSE of the FPM-geo is only 87.25% and 60.01% of those for the TGCN and SVR models, respectively.

The spatial distribution of MAPE implies a negative correlation between prediction error and crowd flow. Further investigation showed that the MAPE of all three models decreased with increasing crowd flow (Fig. 14). However, FPM-geo outperforms the other two methods in both absolute prediction error and robustness of good performance with decreasing crowd flow. For example, when the crowd flow changes from the [3000, 3200] range to the [0, 200] range, the MAPE of FPM-geo increases from 0.009 to 0.069, TGCN increases from 0.01 to 0.074, and SVR increases from 0.015 to 0.092. The reduction in crowd flow of nearly 3000 resulted in decreases of 0.06, 0.063 and 0.077 for the FPM-geo, TGCN and SVR models, respectively. FPM-geo exhibits prediction performance losses of 94.69% and 77.46% due to the same crowd flow reduction for the TGCN and SVR models. When the crowd flow is lower than 200, the RMSE of PPMFPM-geo is only 93.35% and 74.54% of the TGCN and SVR models, respectively.

Time distribution of the prediction error. The time distributions of the two prediction errors are shown in Figs. 15, 16, 17, 18 and 19. The RMSE, MAE, MAPE, WMAE, and WMAPE of the TGCN and SVR models exhibit three peaks in the morning rush hours (7:00–9:00), evening rush hours (17:00–19:00) and the fortnight (21:00–22:00). The proposed FPM-geo has an obviously lower RMSE and only one small peak during 18:00–19:00, which indicates its overall effectiveness in crowd flow prediction. Specifically, during the morning peak, the RMSE of the FPM-geo remains below 15, which is over 22% lower than that of TGCN and over 47% lower than that of SVR. In the evening peak, the RMSE of the FPM-geo remains below 19, which is over 12% lower than that of the TGCN and over 34% lower than that of SVR.

In general, crowd flow prediction during the daytime exhibits higher error than that during the night. We believe that the underlying reason points to the diversity of travel during the daytime. For example, during the morning and evening rush hours, commuting travel contributes the most to the crowd flow. Citizens from suburban areas travel to downtown areas. The impacts of the crowd flow from distant areas increase considerably. The FPM-geo can better address the spatial and temporal features than the SVR and TGCN and outperforms

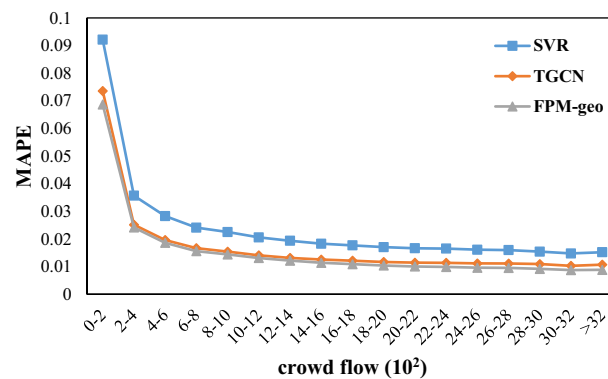


Figure 14. MAPE distributions based on the number of people in the grid.

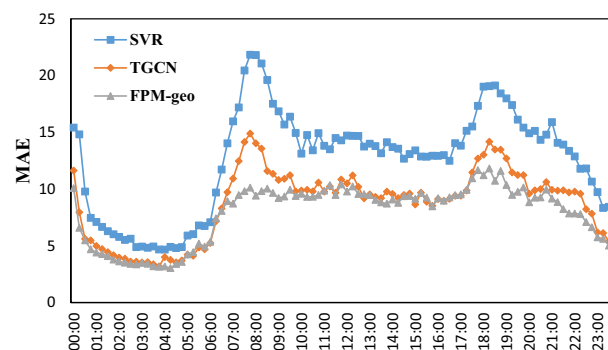


Figure 15. Time distribution of the MAE of each model.

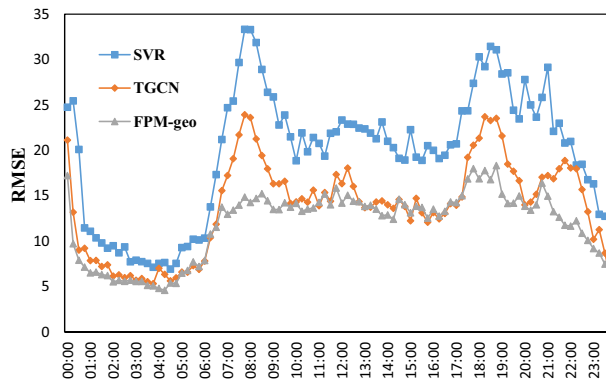


Figure 16. Time distribution of the RMSE of each model.

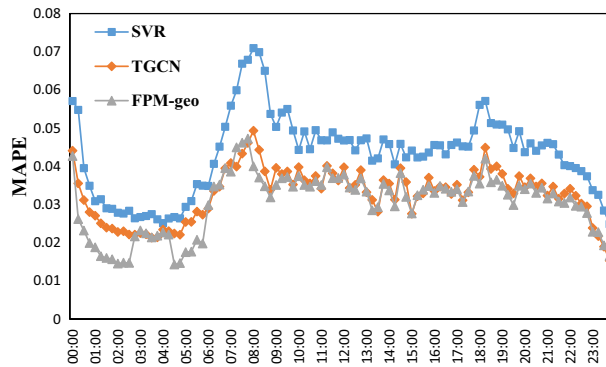


Figure 17. Time distribution of the MAPE of each model.

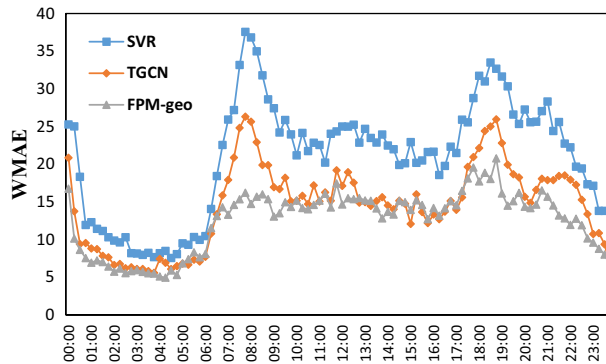


Figure 18. Time distribution of the WMAE of each model.

the other two models in crowd flow prediction. During the night, especially from 1:00 to 6:00, the travel volume remained at a very low level. The impacts from other regions are small, and the improvement of the proposed FPM-geo is also not strong.

Conclusion and discussion

In this paper, we propose an urban crowd flow spatial distribution prediction model by integrating multiple geographic characteristics. The residual multigraph convolution network is applied to integrate the proximity, functional similarity and connectivity relationships that affect the crowd flow prediction effectiveness. LSTM is applied to model the temporal features of the local crowd flow dynamics. Four days of data with more than 0.3 million users validated the effectiveness of the proposed FPM-geo by comparative analysis with six typical methods in existing studies. The model proposed in this paper integrates three geographic features to improve the performance of urban crowd flow prediction. We analyze the impact of several geographical features on the

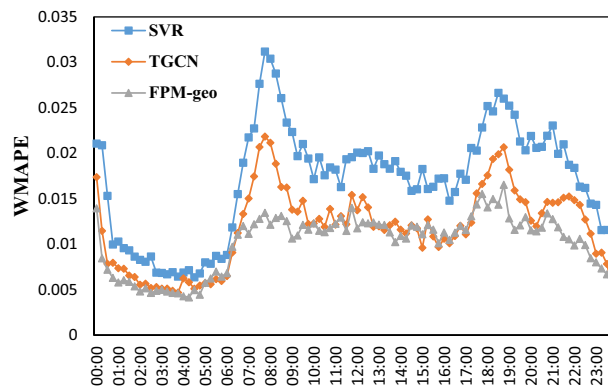


Figure 19. Time distribution of the WMAPE of each model.

prediction accuracy from different perspectives and provide some new insights for related research on urban crowd flow prediction.

First, the proposed method can make a contribution to applications based on the prediction of crowd flows due to its performance in crowd flow prediction in both the temporal dimension and the spatial dimension. In fact, rapid changes in crowd flow in space and time can create problems in public safety, response to extreme climate events, epidemic control, and traffic management, etc. For example, stampedes occurred during the Hajj pilgrimage to Mecca in Saudi Arabia in 2015 and during the Shanghai Bund. Recently, stampedes occurred in Seoul, South Korea. If the crowd flow can be predicted in advance, protection policies could be implemented earlier. We can prevent such catastrophic events from happening or reduce their probability. The prediction of the crowd flow can also contribute to higher urban transportation efficiency. The travel demand to support the vehicle scheduling optimization operation is usually derived from the history travel log. Due to multiple travel modes or various travel service companies, related data can hardly capture the overall travel demand appropriately. The estimated travel demand will reach the upper bound limited by the current service supplies. However, the crowd prediction method can provide a better estimation for the travel demand because it reflects the amount of users.

Second, the effectiveness of the three factors and the spatial and temporal patterns of the prediction performance have high potential generalization in other areas with similar spatial distribution patterns in terms of road networks and urban functions. In general, Xining exhibits “single center” patterns in its spatial distribution of urban functions and road networks. The major government departments, the best hospitals and the most popular shopping malls are all concentrated around a cross area of “two developed axes” along rivers across the city. The road network density and the population density are both the highest in the center area and gradually decrease when moving toward the suburbs. The distribution patterns of the above features are common in cities in China (Beijing, Shanghai, Chengdu, etc.), as well as in the world (e.g., the London, Paris, Tokyo). To better reveal the effectiveness of the selected factors, three typical regions (i.e., center area, suburbs and the outer suburbs) have been selected to compare the contribution of different factors. Therefore, we believe that the proposed model can still work and the three spatiotemporal factors can also contribute a positive impact on improving the crowd prediction performance.

There are still some shortcomings in this paper that need further research. First, the effectiveness of the model on weekends or holidays has not been tested due to dataset limitations. Considering the higher diversity of weekend travel patterns, we believe that the proposed model can achieve better performance than the model chosen in this study. Second, due to the data, we only considered three geographical features. If other urban data (such as social life data, health data, etc.) can be collected in the future, we can better understand the relationship between the urban environment, human activities and complex interactions. Third, the potential overfitting issue of the proposed model requires further testing due to the limited time period and spatial extent of the dataset. It will be promising and valuable to test the performance of the proposed model for other cities.

Data availability

The data presented in this study are available on request from the corresponding author.

Received: 7 July 2022; Accepted: 27 January 2023

Published online: 30 January 2023

References

1. Deville, P. *et al.* Dynamic population mapping using mobile phone data. *Proc. Natl. Acad. Sci.* **111**, 15888–15893 (2014).
2. Wardrop, N. *et al.* Spatially disaggregated population estimates in the absence of national population and housing census data. *Proc. Natl. Acad. Sci.* **115**, 3529–3537 (2018).
3. Vlahogianni, E. I., Golias, J. C. & Karlaftis, M. G. Short-term traffic forecasting: Overview of objectives and methods. *Transp. Rev.* **24**, 533–557 (2004).
4. Chandra, S. R. & Al-Deek, H. Predictions of freeway traffic speeds and volumes using vector autoregressive models. *J. Intell. Transp. Syst.* **13**, 53–72 (2009).
5. Yu, H., Wu, Z., Wang, S., Wang, Y. & Ma, X. Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors* **17**, 1501 (2017).

6. Zhao, L. *et al.* T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* **21**, 3848–3858 (2019).
7. Zhang, J., Zheng, Y. & Qi, D. in *Thirty-First AAAI Conference on Artificial Intelligence*.
8. Pappalardo, L. *et al.* An analytical framework to nowcast well-being using mobile phone data. *Int. J. Data Sci. Anal.* **2**, 75–92 (2016).
9. Jin, W., Lin, Y., Wu, Z. & Wan, H. in *Proceedings of the 2nd International Conference on Compute and Data Analysis* 28–35.
10. Hamed, M. M., Al-Masaeid, H. R. & Said, Z. M. B. Short-term prediction of traffic volume in urban arterials. *J. Transp. Eng.* **121**, 249–254 (1995).
11. Bossanyi, E. Short-term wind prediction using Kalman filters. *Wind Eng.* **9**, 1–8 (1985).
12. Ahmed, M. S. & Cook, A. R. *Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques* (1979).
13. Smith, B. L. & Demetsky, M. J. Traffic flow forecasting: Comparison of Modeling Approaches. *J. Transp. Eng.* **123**, 261–266 (1997).
14. Smith, B. L., Williams, B. M. & Oswald, R. K. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transp. Res. Part C: Emerg. Technol.* **10**, 303–321 (2002).
15. Sun, S., Zhang, C. & Yu, G. A Bayesian network approach to traffic flow forecasting. *IEEE Trans. Intell. Transp. Syst.* **7**, 124–132 (2006).
16. Smola, A. J. & Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **14**, 199–222 (2004).
17. Liu, Q., Wu, S., Wang, L. & Tan, T. in *Thirtieth AAAI Conference on Artificial Intelligence*.
18. Zonoozi, A., Kim, J.-j., Li, X.-L. & Cong, G. in *IJCAI 3732–3738*.
19. Moravčík, M. *et al.* Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science* **356**, 508–513 (2017).
20. Lv, Y., Duan, Y., Kang, W., Li, Z. & Wang, F.-Y. Traffic flow prediction with big data: a deep learning approach. *IEEE Trans. Intell. Transp. Syst.* **16**, 865–873 (2014).
21. Park, D. & Rilett, L. R. Forecasting freeway link travel times with a multilayer feedforward neural network. *Comput.-Aided Civ. Infrastruct. Eng.* **14**, 357–367 (1999).
22. Fu, R., Zhang, Z. & Li, L. in *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)* 324–328 (IEEE).
23. Cao, X. *et al.* Interactive temporal recurrent convolution network for traffic prediction in data centers. *IEEE Access* **6**, 5276–5289 (2017).
24. Liu, Y., Zheng, H., Feng, X. & Chen, Z. in *2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)* 1–6 (IEEE).
25. Ma, X. *et al.* Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. *Sensors* **17**, 818 (2017).
26. Chen, J. *et al.* Fine-grained prediction of urban population using mobile phone location data. *Int. J. Geogr. Inf. Sci.* **32**, 1770–1786 (2018).
27. Zhang, G. *et al.* Large-scale, fine-grained, spatial, and temporal analysis, and prediction of mobile phone users' distributions based upon a convolution long short-term model. *Sensors* **19**, 2156 (2019).
28. Wu, Y. & Tan, H. Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework. arXiv preprint [arXiv:1612.01022](https://arxiv.org/abs/1612.01022) (2016).
29. Welling, M. & Kipf, T. N. in *International Conference on Learning Representations (ICLR 2017)*.
30. Chai, D., Wang, L. & Yang, Q. in *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* 397–400.
31. Sun, J. *et al.* Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks. *IEEE Trans. Knowl. Data Eng.* **6**, 66 (2020).
32. Tobler, W. R. A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **46**, 234–240 (1970).
33. Zhu, A. X., Lu, G., Liu, J., Qin, C. Z. & Zhou, C. Spatial prediction based on Third Law of Geography. *Ann. GIS* **24**, 225–240 (2018).
34. Anselin, L. *What is Special About Spatial Data? Alternative Perspectives on Spatial Data Analysis* (89–4) (1989).
35. Defferrard, M., Bresson, X. & Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. *Adv. Neural Inf. Process. Syst.* **29**, 66 (2016).
36. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
37. Alin, A. Simpson's paradox. *Wiley Interdiscip. Rev.: Comput. Stat.* **2**, 247–250 (2010).
38. Zhao, Z. *et al.* The effect of temporal sampling intervals on typical human mobility indicators obtained from mobile phone location data. *Int. J. Geogr. Inf. Sci.* **33**, 1471–1495 (2019).
39. Geng, X. *et al.* in *Proceedings of the AAAI Conference on Artificial Intelligence* 3656–3663.
40. Giannotti, F. *et al.* Unveiling the complexity of human mobility by querying and mining massive trajectory data. *VLDB J.* **20**, 695–719 (2011).
41. Xu, Y. *et al.* Another tale of two cities: Understanding human activity space using actively tracked cellphone location data. *Ann. Am. Assoc. Geogr.* **106**, 489–502 (2016).

Author contributions

Data curation, X.Y.; Conceptualization, Methodology, Writing-Original Draft Preparation, Y.Z., S.W. and Z.Z.; writing-review and editing Z.Z. and Z.F. All authors have read and agreed to the published version of the manuscript.

Funding

This research was supported by the National Natural Science Foundation of China (No. 42201500, 42271468, 41771473); the Fujian Cooperation Project between Universities and Enterprises, Fujian, China (No. 2021H6004); the Founding of the Key Lab of Spatial Data Mining & Information Sharing, Ministry of Education (No.2022LSDMIS03); Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources (KF-2022-07-005).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-29000-5>.

Correspondence and requests for materials should be addressed to Z.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023