International Journal of Digital Earth

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tjde20

Self-paced Gaussian-based graph convolutional network: predicting travel flow and unravelling spatial interactions through GPS trajectory data

Shuhui Gong, Jialong Liu, Yuchen Yang, Jingyi Cai, Gaoran Xu, Rui Cao, Changfeng Jing & Yu Liu

To cite this article: Shuhui Gong, Jialong Liu, Yuchen Yang, Jingyi Cai, Gaoran Xu, Rui Cao, Changfeng Jing & Yu Liu (2024) Self-paced Gaussian-based graph convolutional network: predicting travel flow and unravelling spatial interactions through GPS trajectory data, International Journal of Digital Earth, 17:1, 2353123, DOI: 10.1080/17538947.2024.2353123

To link to this article: https://doi.org/10.1080/17538947.2024.2353123

Published online: 20 May 2024.

INTERNATIONAL JOURNAL OF DIGITAL EARTH
2024, VOL. 17, NO. 1, 2353123
https://doi.org/10.1080/17538947.2024.2353123

Self-paced Gaussian-based graph convolutional network: predicting travel flow and unravelling spatial interactions through GPS trajectory data

Shuhui Gong[a], Jialong Liu[a], Yuchen Yang[a], Jingyi Cai[b], Gaoran Xu[a], Rui Cao [c], Changfeng Jing [a] and Yu Liu [d]

[a]School of Information Engineering, China University of Geosciences, Beijing, People's Republic of China; [b]School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, People's Republic of China; [c]Department of Land Surveying and Geo-Informatics & Otto Poon Charitable Foundation Smart Cities Research Institute, The Hong Kong Polytechnic University, Hong Kong SAR, People's Republic of China; [d]Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing, People's Republic of China

ABSTRACT

Spatial interaction research is particularly important for geographical analyses, as it plays a crucial role in extracting travel patterns. However, previous studies on spatial interactions have not adequately considered regional population variations over time, resulting in insufficiently precise travel predictions. Moreover, the threshold of spatial correlations is difficult to determine. Existing studies have assumed fully connected spatial correlation matrices, which is not realistic. To address these limitations, we proposed the Self-paced Gaussian-Based Graph Convolutional Network (SG-GCN) to automatically estimate the threshold of spatial correlations for travel flow predictions. It incorporates a temporal dimension into spatial relationship matrices to enhance the accuracy of vehicle flow predictions. In particular, Gaussian-based GCN identifies patterns in a time series of regional flows, enabling more precise capturing of spatial relationships while fusing node and edge features. Building on this model, self-paced contrastive learning automatically sets thresholds to determine the presence or absence of spatial relationships. The model's performance was verified through two empirical case studies conducted in New York City, USA, and Ningbo, China, using 2.8 million bicycle-sharing records and 1.25 million taxi trip records, respectively. The proposed model helps delineate mobility patterns in cities of varying scales and with different modes of transportation.

ARTICLE HISTORY
Received 19 October 2023
Accepted 3 May 2024

KEYWORDS
Spatial interaction; travel flow prediction; self-paced contrastive learning; Gaussian process regression; graph convolution network

1. Introduction

Due to the rapid increase in urban populations, spatial interactions in human geography have attracted considerable attention from businesses and academics (Fotheringham 1981; Roy and Thill 2003). Studies of human flows have covered a variety of areas, including human migrations (Sîrbu et al. 2021), changes in urban environments (Gonzalez, Hidalgo, and Barabasi 2008), natural disaster prediction (Gray and Mueller 2012), and the spread of epidemics (Kraemer et al. 2020).

CONTACT Changfeng Jing jingcf@cugb.edu.cn School of Information Engineering, China University of Geosciences, Beijing, People's Republic of China

Travel flow prediction is particularly important for urban and transportation planning (Gong et al. 2023). Recent research has shown that travel flow prediction models that consider spatial interactions yield more accurate results (Gong et al. 2022). Specifically, flow data can be collected to explore the intensity of vehicle flows and the connections between entities, which is important for analysing the spatial interaction patterns of different geographical locations (Guo et al. 2012). The precise depiction of dynamic flows within intricate geographical structures is crucial for predicting travel flows, thereby contributing to socio-economic development.

To gain a deeper understanding of urban structures and built-up urban environments (Huang et al. 2022; Liu et al. 2015), various approaches to capturing spatio-temporal features, ranging from statistical to deep learning methods, have been proposed. For instance, Wang et al. (2022) obtained location-related spatial interactions through bidirectional messaging and introduced the Traffic Gated Graph Neural Network (Traffic-GGNN) for real-time-fused spatio-temporal representation modelling to make predictions. Koesdwiady, Soua, and Karray (2016) proposed a deep belief network for traffic and weather forecasts using travel flow data to make more effective predictions of interactions in urban spaces.

However, although ameliorated methods for capturing spatial interactions have been explored, few studies have investigated the temporal correlations between the spatial flows of different departure and destination categories. Previous studies have often calculated spatial and temporal relationships separately, presumably due to the simplicity of this strategy. However, temporal relationships can affect spatial relationships. For example, the interaction patterns during rush hours differ considerably from those during other times of the day. Therefore, it is important to examine how to extract dynamic spatial features in a monitoring network to reflect cyclical patterns and their main temporal variation trends. Moreover, previous studies have tended to assume that all nodes are fully connected. However, in real life, some areas, especially in metropolises, may have very weak or no spatial correlations. Therefore, it is necessary to determine connectivity using a reasonable threshold.

Gaussian process regression (GPR) provides a flexible and comprehensible methodology for examining and forecasting precipitation (Williams and Rasmussen 1995). It offers a principled quantification of uncertainty by design. Specific prediction intervals can be derived from the posterior predictive distribution. When accompanied by robust uncertainty estimates, predictions can be pivotal to downstream decision-making. GPR also offers a flexible prior of functions, allowing the encoding of specific inductive biases through kernel construction. For instance, properties such as periodicity, smoothness, and spatial heterogeneity can be encoded through careful specification of the covariance kernel. Since assessing the quality of embedded graphs relies on a priori knowledge of the downstream prediction task and the graph structure, Self-Paced Graph Contrastive Learning (SPGCL) (Li et al. 2022) has been proposed to compensate for this aspect, with its adjacency matrix providing filterable thresholds for information between nodes. Specifically, the values of more weakly correlated positions are set to 0 according to the thresholds, and the correlations are extracted directly from the data.

To address the above-mentioned key issues and to fill the extant research gaps, we propose the Self-paced Gaussian-Based Graph Convolutional Network (SG-GCN) for spatio-temporal correlation extraction from GPS trajectories. This framework employs non-parametric GPR, graph convolutional networks (GCNs), and gated recurrent units (GRUs) and can predict spatio-temporal travel volumes based on interregional relationships. The main contributions of this work are as follows:

- GPR is innovatively utilised to extract spatial correlations through the temporal variations of travel flows, rather than investigating these two aspects separately, thereby achieving higher travel flow prediction accuracy than that achieved in previous studies. This non-parametric model captures node features based on their similarity, enabling the combination of various pieces of prior information and uncertainty to learn spatio-temporal features at high latitudes without relying on linear two-dimensional combinations. Therefore, our framework represents spatial correlations more comprehensively and accurately compared to previous research.

- The model also incorporates the recently proposed SPGCL to further estimate the connectivity of the nodes of the graph structure, largely filtering out unnecessary information and improving computational efficiency. SPGCL investigates spatial correlations in the temporal dimension, providing a novel approach to computing the spatio-temporal interactions between nodes in the absence of predefined graphs. By increasing the margin of differentiation between positive and negative neighbours to learn node linkages, SPGCL can produce optimal graphs using a self-paced strategy, which provides a threshold for the spatial correlation matrix generated by GPR.

We transformed point-of-interest (POI) data with geographical attributes into feature vectors combined with travel flows, considering multidimensional features in temporal variations, through which we distinguished the different attributes of the locations involved and explored various crowd choices under the influence of the environment. To test the model's performance, we conducted two case studies in New York City, USA, and Ningbo, China, using 2.8 million bicycle-sharing records and 1.25 million taxi ride records, respectively. We also explored residents' travel preferences in both cities.

The rest of this paper is organised as follows: Section 2 presents previous research on spatial interactions in travel flow predictions. Section 3 describes the theories and methodology used in this study. Sections 4 and 5 present the case studies of New York City and Ningbo, respectively, and the evaluation of the prediction results through test data. Section 6 concludes the paper.

2. Literature review

2.1. Travel flow prediction

Both parametric and non-parametric models have been used for travel flow predictions (Medina-Salgado et al. 2022). Parametric methods include traditional statistical methods and machine learning approaches. Traditional statistical methods, such as regression models and autoregressive integrated moving averages, are based on assumptions, which are simpler and easier to interpret (Ahmed and Cook 1979; Russell 2010; Shan, Zhao, and Xia 2013; Williams and Hoel 2003). However, the variables in regression analyses are highly correlated and non-linear, which makes the non-linear functions of a model particularly difficult to define (Smith and Demetsky 1994). Therefore, such methods have difficulty representing spatial relationships, and their prediction accuracy is generally low. In recent years, various machine learning methods have been used for data mining and predictions, such as the improved k-nearest neighbours (KNN) algorithm (Zhang et al. 2013), support vector machines (SVMs) (Feng et al. 2018), and Bayesian algorithm (Zheng, Lee, and Shi 2006). Lin, Lin, and Gu (2022) introduced a model for screening spatial delayed traffic sequences based on the maximum information coefficient. The screened sequences are transformed into traffic state vectors, and the traffic flow is forecast using a combination of support vector regression and KNN. However, parametric methods often lack the travel flow prediction accuracy of non-parametric methods.

Among non-parametric methods, neural networks have been widely used with good results (Wang et al. 2022), sometimes showing superiority to SVMs and KNN. Recurrent neural networks and optimised LSTM networks have successfully performed long- and short-term sequential learning tasks, and hybrid neural network approaches have been used in various fields Sutskever, Vinyals, and Le (2014). For instance, Zhang, Zheng, and Qi (2017) proposed a convolution-based spatio-temporal residual network to make predictions about a specific spatio-temporal. Zheng, Lee, and Shi (2006) employed an adaptive Bayesian–neural network approach to reduce prediction errors using a single radial basis function (RBF) or BP, demonstrating the considerable potential of this composite model. Chen et al. (2020) smoothed loop data through wavelet filtering, moving average modelling, and Butterworth filtering and introduced an ANN to predict traffic flows for different time spans. GCNs are also widely used non-parametric models (Zhao et al. 2020). Zhao et al. (2020) proposed the temporal graph convolutional network (T-GCN) to predict spatio-temporal flows, demonstrating the ability of

GCNs to extract complex graph structure data and achieve high prediction accuracy. Recently, various self-paced spatio-temporal learning approaches have also been used to address challenges related to graph networks. For example, Li et al. (2022) proposed the SPGCL model, which not only captures temporal and spatial dependencies but also learns information linkages by optimising the difference between positive and negative neighbours, thus improving the network structure. Zhang et al. (2023) introduced the Automated Spatio-Temporal (AutoST) graph contrastive learning framework based on heterogeneous graph neural architecture, which captures multi-view region dependencies in POI semantics, mobility flow patterns, and locations.

Although the approaches described above can predict spatial travel flows with some accuracy, standard models do not account for the social characteristics of crowd flows, which may result in the loss of genuine information (Zhou et al. 2020). Moreover, although previous studies have used spatial correlations to make predictions, they have not sufficiently evaluated them over time, instead simply focussing on the total statistics of geographical relationships over a given period, which may result in less accurate predictions.

2.2. Spatial interaction analysis

Previous research has used various criteria to measure spatial interactions, mainly geographical distances, virtual network flows, and travel mobility flows. Geographical distances have been the most commonly used to represent spatial proximity since the closer together objects are, the more closely they interact with one another (Fotheringham 1981; Miller 2004). The potential of using big data to quantify geographical correlations increases as information networks grow, and community probes from online networks are widely used to learn spatial relationships (Cheng et al. 2011; Liu et al. 2014). For example, Liu et al. (2014) used social media data to identify intra-town tourism patterns. Chi et al. (2016) used mobile phone call data to explore interregional characteristics and structures. Chen et al. (2022) employed advanced artificial intelligence modelling and computer vision techniques to collect motion data and analyse vehicle spatial interactions. However, while these studies have provided some summaries of inter-location features, distance data alone cannot provide sufficient information on spatial qualities, and online data may not be representative of real human behaviour due to the limitations of their audience (Eysenbach and Wyatt 2002).

Travel flow data constitute a richer and more accessible source of information for interpreting the spatial interactions of crowds (Liu, Gao, and Lu 2019; Tu et al. 2018). To compensate for data scarcity in emerging cities, many recent studies have combined a priori knowledge with scenario-based data-driven approaches (Kong et al. 2023). Lin, He, and Peeta (2018) presented an updated graph convolutional neural network (GCNN-DDGF) to reveal hidden heterogeneous pairwise correlations across stations and anticipate dynamic demand at the station level in a large-scale urban bicycle-sharing network. Using a deep learning T-GCN framework, Gong et al. (2023) performed a linear superposition of the scene attributes of parking behaviour combined with temporal features, optimising the accuracy of spatio-temporal prediction to a certain extent. Kong et al. (2022) used a multimodal technique to predict bus passenger flows using graph networks and a deep clustering approach to discover hidden passenger interaction patterns. While these studies have contributed to the temporal representations of scene semantics, some have been less attentive to capturing continuous dynamic changes or have failed to achieve non-linear learning of complex geographical features.

3. Methodology

3.1. Overview of proposed framework

In this study, we propose the SG-GCN to accurately predict travel flows and extract spatial correlations between areas. The overall structure of the proposed framework is depicted in Figure 1. We used two types of data: travel origin–destination (OD) and POI data. The format of the OD

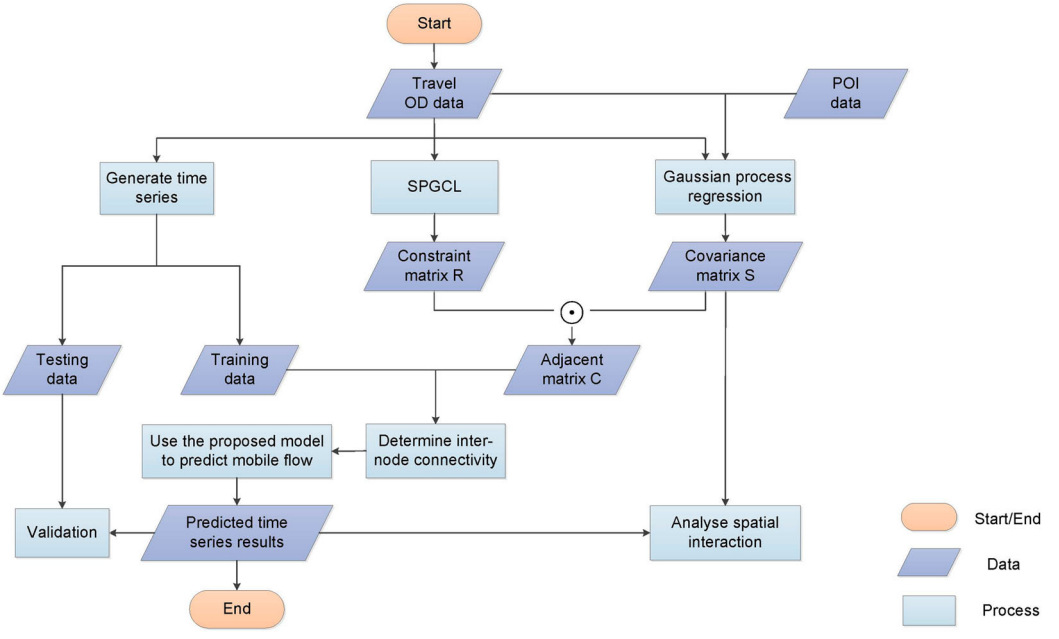

Figure 1. The overview of the framework.

Table 1. The format of travel OD data.

ID	D_t	D_{sx}	D_{sy}	D_{ex}	D_{ey}
159	01/04/2023 17:24:00	40.684	−73.989	40.722	−74.011

data is shown in Table 1. They include departure time D_t, departure latitude D_{sx}, departure longitude D_{sy}, ending longitude D_{ex}, and ending latitude D_{ey}. We divided the research area into 1 km × 1 km squares and generated a travel time series for each square from the OD data. We then fed the time series and POI data of each square into GPR to extract features while simultaneously generating the distance matrix between sites. Specifically, we created four types of adjacency matrices: (i) a distance matrix (A), (ii) a covariance matrix of time series generated by GPR (S'), (iii) a covariance matrix built by GPR using the time series and POI data (S), and (iv) a linear combination of matrices (i) and (iii) (C'). We used the SPGCL approach to determine the connectivity of the nodes in the matrix. We then fed the adjusted adjacent matrix C and time series data into the GCN and GRU, respectively, to predict travel flows. The algorithm process is presented in Figure 2. Finally, we compared the performance of the SG-GCN and each matrix by validating the results. The details of each step are presented below.

3.2. Time series generation

Algorithm 1 explains the process of generating a time series. The input travel flow dataset D indicates the amount of outflow from a specific area. B represents the prediction of node sequences, including the predicted node longitude $B.x$ and latitude $B.y$ and the number of predicted nodes $B.l$. Each day, the travel data of each block are traversed by one hour, and when the i-th time and the coordinates of the j-th block are identified, the $D^*_{i,j}$ value plus one is obtained. Finally, D^* returns a horizontal time coordinate, a vertical block number coordinate, and the travel value.

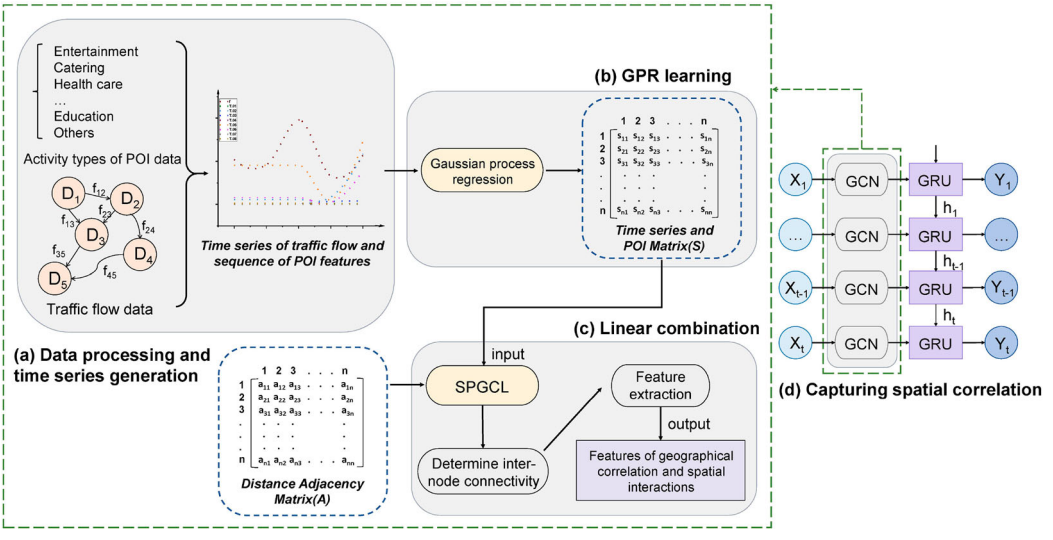

Figure 2. The process of SG-GCN, which takes matrix *C* as an example.

Algorithm 1: Process travel flow into a time series

Data: *D*, *B*, *T*
// *D*:Travel flow data, *B*:latitude and longitude information of blocks, *T* :Time Series
Result: D^*
// Time series data D^*
D^*initialized as a zero matrix
$T.l = T.length()$ // the length of *T*
the length of *T* **foreach** *element of the* D **do**
  for $i \leftarrow 0$ ***to*** T.l **do**
    if $D_t = T_i$ **then**
      for $j \leftarrow 0$ ***to*** B.l **do**
        if $D_x = B_j.x$ *and* $D_y = B_j.y$ **then**
          $D^*_{ij} += 1$
        end
      end
    end
  end
end

3.3. Spatial correlation modelling

As mentioned earlier, four types of adjacent matrices are used. Figure 2 shows the generated process of C', where $D.i$ denotes the i-th district, $t_{i,j}$ represents the number of activities in the j-th category in $D.i$, $f_{i,j}$ represents the outflow from $D.i$ to $D.j$ in a certain time, $a_{i,j}$ indicates the distance from $D.i$ to $D.j$, and $s_{i,j}$ indicates the similarity between regions.

3.3.1. Distance matrix

As previously noted, geographical distances are important for studying spatial interactions. Traditional distance calculation methods include the Euclidean, Manhattan, and Chebyshev distances, which can be computed geometrically using a road's coordinates. The distance matrix is generated by calculating sophisticated distances derived by merging actual road information with binding roads. The latitude and longitude of the i-th and j-th blocks' central points B are entered into a web API, and the output data provide the real driving distance between the two blocks. We tested

various distance attenuation coefficients β, ranging from -2 to 0, and selected to result with the highest accuracy. We regarded the computed distance matrix as the adjacency matrix A after threshold filtering. The process is shown in Algorithm 2.

Algorithm 2: Distance matrix preprocessing

Input: B, p
// B:latitude and longitude of the block centre points, p:The filtering percentage
Output: A
// A:Adjacency matrix
$a_{i,j} = API(B_i, B_j)$ // The driving distance between block i and block j
$a' = a.sort()$ // Sort the distances between blocks in descending order
$n = a'.lenght()$ // The length of a'
foreach *element* $a_{i,j}$ *of the* a **do**
  if $a_{i,j} <= a'_{n*\beta}$ **then**
    $A_{i,j} = a^{\beta}_{i,j}$
  end
  else
    $A_{i,j} = 0$
  end
end

3.3.2. Time series matrix

The covariance matrix is generated by GPR, which is a non-parametric regression technique based on Bayesian probability theory. GPR can estimate confidence intervals for the expected outcomes and model non-linear connections between continuous variables. It considers the target variable in the context of a set of random variables. The joint distribution of any subset of this set of variables is normal, and it is fully specified by the mean function $\mu(x)$ and the covariance function $k(x, x')$, with the former describing the mean GP trend and the latter controlling the function's inductive bias and support under the prior:

$$g \sim GP(\mu, k) \tag{1}$$

where the representation function f is described as a GP with a mean value function μ and a covariance function k. Suppose there is a dataset D containing the input $x = \{x_1, x_2, x_3, \ldots, x_n\}$ and the corresponding output $y = \{y_1, y_2, y_3, \ldots, y_n\}$

$$D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \tag{2}$$

Given a training dataset D and a new test point x, we wish to predict its corresponding output y and assume that y and $f(x)$ follow a joint Gaussian distribution:

$$\begin{bmatrix} g(\mathbf{x}) \\ \mathbf{y}^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_g \\ \mu_y \end{bmatrix}, \begin{bmatrix} K_{gg} & K_{gy} \\ K_{gy}^T & K_{yy} \end{bmatrix}\right) \tag{3}$$

where $K_{gg} = k(x, x)$, $K_{gy} = k(x, x^*)$, *and* $K_{yy} = k(x^*, x^*)$. Based on dramatic Bayes' theorem, we can calculate the posterior distribution $p(g(x^*)|D)$, i.e. the output probability distribution of the prediction point x for a known dataset D. This posterior distribution is also a Gaussian distribution whose mean and variance can be calculated based on the information in the training dataset D. The process is expressed by Equation (4):

$$f \sim \mathcal{N}\left(K_{gy}^T K_{gg}^{-1}\mathbf{y} + \mu_g, K_{yy} - K_{gy}^T K_{gg}^{-1} K_{gy}\right). \tag{4}$$

The model selection problem in the GP refers to the selection of the covariance function (kernel function) and its hyperparameters (Rasmussen 2004). The most commonly used kernels in machine learning include polynomial kernels, RBFs, and automatic relevance determination

(ARD). The ARD framework acts by reducing the inverse length scale to prune irrelevant dimensions. The SE-ARD framework, a translation-invariant stationary kernel that delivers endlessly smooth and differentiable samples in the function space, is one of the most commonly used kernel function frameworks in GPR applications. The SE-ARD kernel can be defined as:

$$k(\mathbf{x}, \mathbf{x}')_{\mathrm{SE-ARD}} = \sigma_g^2 \exp\left\{-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_d - x_d')^2}{\ell_d^2}\right\} \tag{5}$$

where x denotes the observed predictors, $\sigma_g^2(\cdot)$ is the variance of the kernel function, x_d and x_d' are the values of the inputs x and x', respectively, in the d-th feature, and ℓ_d refers to a characteristic length representing a distance in the input space and indicating the reference capability of the input variable.

Since a single SE-ARD kernel cannot meet the multiple requirements of travel flow prediction, Lalchand et al. (2022) developed a spatio-temporal kernel that multiplies both to superimpose the temporal and spatial components (Equation (6)) and handle the temporal characteristics using the PER kernel (Equation (7)):

$$\begin{aligned} k_{\mathrm{st}}(x_i, x_j) = & \underbrace{k_{\mathrm{SE-ARD}}\left(\left(x_i^{\mathrm{lat}}, x_i^{\mathrm{lon}}\right), \left(x_j^{\mathrm{lat}}, x_j^{\mathrm{lon}}\right)\right) \times k_{\mathrm{PER}}\left(x_i^t, x_j^t\right)}_{\text{temporal}} \\ & + \underbrace{k_{\mathrm{SE-ARD}}\left(\left(x_i^{\mathrm{lat}}, x_i^{\mathrm{lon}}\right), \left(x_j^{\mathrm{lat}}, x_j^{\mathrm{lon}}\right)\right)}_{\text{spatial}} \end{aligned} \tag{6}$$

$$k_{PER}(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left\{-\frac{2\sin^2\left(\pi|\mathbf{x}_i - \mathbf{x}_j|/P\right)}{\ell^2}\right\} \tag{7}$$

where (x_i^{lat}, x_i^{lon}) denotes the latitude and longitude of the i-th node, (x_i^t, x_j^t) denotes the time of the i-th and j-th nodes, and P is the periodicity parameter.

The GP is a non-parametric model that does not require parameters for the training model and can be uniquely calculated once the kernel function and training data have been obtained. However, The kernel function itself has hyperparameters, such as ℓ and σ^2, which alter the measure of similarity between sample points, thereby changing the overall probability distribution function.

We optimised the hyperparameters using Bayesian optimisation (BayesOpt) to maximise the marginal log-likelihood (MLL). The optimisation process involved iteratively computing the gradients of the MLL with respect to the hyperparameters σ and ℓ to find the optimal combination that maximised the likelihood of the observed data y. BayesOpt is a probabilistic method for optimising objective functions based on Bayes' theorem, as described by Rasmussen and Williams (2006). The MLL computation method used in this study was:

$$\log p(y \mid \sigma, \ell) = \log N\left(0, K_{yy}(\sigma, \ell)\right) = -\frac{1}{2}y^T K_{yy}^{-1} y - \frac{1}{2}\log|K_{yy}| - \frac{N}{2}\log(2\pi) \tag{8}$$

where $|\cdot|$ indicates the determinant of the matrix. By maximising the MLL, we identified the best parameters.

3.3.3. Time series and POI matrix

As travel data are closely associated with daily life, along with these data, we inputted POI data into the GPR module to obtain another covariance matrix. The updated GPR requires the addition of a POI kernel component. The improved kernel function is expressed as:

$$k_{\text{st-poi}}\left(x_i, x_j\right) = \underbrace{k_{\text{SE-ARD}}\left(\left(x_i^{\text{lat}}, x_i^{\text{lon}}\right), \left(x_j^{\text{lat}}, x_j^{\text{lon}}\right)\right) \times k_{\text{PER}}\left(x_i^t, x_j^t\right)}_{\text{temporal}} + \underbrace{k_{\text{SE-ARD}}\left(\left(x_i^{\text{lat}}, x_i^{\text{lon}}\right), \left(x_j^{\text{lat}}, x_j^{\text{lon}}\right)\right)}_{\text{spatial}} + \underbrace{k_{\text{SE-ARD}}\left(x_i^{poi}, x_j^{poi}\right)}_{\text{poi}} \tag{9}$$

where x^{poi} and x^{poi} represent the POI features combined with flow data.

3.3.4. Linear combination matrix

Inspired by previous work, we used a linear combination of both factors to form a new adjacency matrix C' that describes spatial correlations (Gong et al. 2023):

$$C' = a \cdot A + b \cdot S \tag{10}$$

where a denotes the coefficients of the distance matrix A, and b denotes the coefficients of the covariance matrix S computed through GPR based on POI data. The values of a and b describe the impacts of travel distance and similarities on a given region's spatial interactions, respectively. After continuously changing the parameters, we produced the optimal adjacency matrix and fed it into the GCN to obtain the spatial correlations.

3.4. Self-paced graph construction

The connectivity of the graph is determined through self-paced graph construction. The SPGCL method employs messages passing on a graph structure $g = (V, E, O)$, where V represents nodes, E represents edges, and O is the adjacency matrix.

As E is often not explicitly defined, SPGCL initially uses KNN to produce an initial neighbourhood matrix $O^{t,0}$ with low recall but high precision when $k \ll N$. The genuine neighbourhood is inferred through the score function g, the observed traffic flow value x^t, and the position of node V. To better distinguish edges, SPGCL separates all possible edges into three parts: a positive set labelled E^P, a negative set labelled E^N, and an unlabelled set E^U (i.e. $E = E^P \cup E^N \cup E^U$). Potential advantages with self-confidence scores over δ^+ are added to E^P, while potential advantages with scores below δ^- are added to E^N. The process is repeated to obtain $O^{t,K}$ until $E^U = \emptyset$.

SPGCL iteratively searches for the k-true neighbours of a node through maximised interactivity information (MI). To compute the MI without losing generality, SPGCL first establishes a relative spatio-temporal coordinate system with node V_i as the origin and measures the MI using the following relationships between node V_i and its neighbours:

$$w_{i,j}^{dis} = \| \mathbf{V}_i - \mathbf{V}_j \|_2 \tag{11}$$

$$w_{i,j}^{\text{ang}} = \sin\left(\theta_{i,j}\right) = \frac{\mathbf{V}_i[2] - \mathbf{V}_j[2]}{w_{i,j}^{dis}} \tag{12}$$

$$w_{i,j}^{\text{seq}} = DTW\left(\mathbf{x}_i^t, \mathbf{x}_j^t\right) \tag{13}$$

$$w_{i,j}^t = \mathbf{x}_i^t - \mathbf{x}_j^t \tag{14}$$

where $w_{i,j}^{dis}$ and $w_{i,j}^{ang}$ are the spatial polar coordinates of node V_j with respect to V_i, $w_{i,j}^{seq}$ represents the similarity between the time series w_i^t and w_j^t computed by the DTW distance, which has been widely used to measure the similarity between sequential data, and $w_{i,j}^t$ is the sequence distance between w_i^t.

The details of the implementation of SPGCL to calculate connectivity are as follows:

$$\mathbf{w}_{i,j}^{t} = \left[w_{i,j}^{dis}, w_{i,j}^{\text{ang}}, w_{i,j}^{seq}, w_{i,j}^{t} \right] \tag{15}$$

$$\mathbf{r}_{i,j}^{t} = f\left(\mathbf{w}_{i,j}^{t}; \phi\right) = MLP\left(\mathbf{w}_{i,j}^{t}\right) \tag{16}$$

$$\mathbf{r}_{i,c}^{t} = f\left(C_i^k; \Phi\right) = \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \left(\mathbf{r}_{i,j}^{t} \odot \mathbf{w}_{i,j}^{t}\right) \tag{17}$$

where $r_{i,j}^{t}$ is the spatio-temporal embedding and MLP is the multilayer perceptron. Then SPGCL combines the characteristics of the neighbouring nodes to generate the contextual embedding with regard to node V_i, where $\odot$ is the Hadamard product and N_i is the node's neighbour. Negative Euclidean distance is used to measure the proximity for matching score calculated as follows:

$$s_{i,j}^{t} = f\left(\mathbf{r}_{i,c}^{t}, \mathbf{r}_{i,j}^{t}\right) = -\| \mathbf{r}_{i,c}^{t} - \mathbf{r}_{i,j}^{t} \|_2 \tag{18}$$

Subsequently, SPGCL makes a prediction by aggregating the features of the neighbours and calculates the prediction loss as:

$$\hat{\mathbf{X}}_i = \frac{1}{H|\mathcal{N}_i|} \sum_{h=1}^{H} \sum_{j \in \mathcal{N}_i} \left(\mathbf{M}^h \left(\rho\left(\mathbf{R}_{i,j}\mathbf{U}^h\right) \odot \mathbf{X}_j\right)\right) \tag{19}$$

$$\mathcal{L}_{mse}^{i,k} = \| \mathbf{X}_i - \hat{\mathbf{X}}_i \|_2^2 \tag{20}$$

where $R_{i,j}$ is the constraint matrix, X_j denotes the features of node V_j, U and M denote the linear transformations, H is the number of multi-head aggregations, $\rho(x) = LeakyReLU(x)$ is the non-linear activation function, $\odot$ is the Hadamard product, and C is the matrix consisting of the filtered nodes.

3.5. Travel flow prediction

We used T-GCN, which performs better that other methods (Gong et al. 2023; Zhao et al. 2020), to predict travel flows using the extracted spatial and temporal dependencies (see Figure 2). The two-layer GCN structural model proceeds as follows:

$$f(X, C) = \sigma\left(\hat{C}Relu\left(\hat{C}XW_0\right)W_1\right) \tag{21}$$

$$\hat{C} = \widetilde{D}^{-\frac{1}{2}}\widetilde{C}\widetilde{D} \tag{22}$$

$$\widetilde{C} = C + I_N \tag{23}$$

$$\widetilde{D}_{ii} = \sum j\widetilde{C}_{ij} \tag{24}$$

where X represents the feature matrix, $\widetilde{D}$ represents the degree matrix, $\widetilde{C} = C + I_N$ represents the adjacency matrix (previously generated by GPR), W_0 and W_1 represent the weight matrices in the first and second layers, and $Relu(\cdot)$ represents the activation function.

The GRU procedure is as follows:

$$u_t = \sigma(W_u[f(C, X_t), h_{t-1}] + b_u) \tag{25}$$

$$r_t = \sigma(W_r[f(C, X_t), h_{t-1}] + b_r) \tag{26}$$

$$c_t = tanh(W_u[f(C, X_t), (r_t * h_{t-1})] + b_c) \tag{27}$$

$$h_t = u_t * h_{t-1} + (1 - u_t) * c_t \tag{28}$$

where X_t represents the travel flow data of each node at time t, $f(C, X_t)$ is the process of graph convolution, h_t represents the output at time t, c_t represents the cell state at time t, u_t and r_t represent the update and reset gates at time t, respectively, and W and b are the weights and biases, respectively.

To assess the framework's prediction accuracy, the following four criteria are used:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(y_t - \widehat{y_t})^2} \tag{29}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{n}\left|y_t - \widehat{y_t}\right| \tag{30}$$

$$Accuracy = 1 - \frac{\|Y - \hat{Y}\|_F}{\|\hat{Y}\|_F} \tag{31}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_t - \widehat{y_t})^2}{\sum_{i=1}^{n}(y_t - \bar{Y})^2} \tag{32}$$

where y_t and $\widehat{y_t}$ independently reflect the actual and predicted travel flows of the i-th node at time j, N denotes the number of nodes and the number of time samples, y and $\hat{Y}$ represent the sets of y_t and $\widehat{y_t}$, respectively, and $\bar{Y}$ is the mean value of Y. We calculated the forecasting error using the RMSE and MAE, with lower values indicating more accurate forecasts. We employed R^2 to evaluate the consistency of the projected outcomes in representing the real flow, with higher values indicating greater consistency.

4. Case study in New York City, USA: bike sharing trips prediction

4.1. Study area and data

We conducted a case study of New York City to evaluate the framework's functionality. New York is the largest city in the USA, the economic, commercial, and cultural centre of the USA and the world, and one of the world's top three financial centres.

4.1.1. Travel flow data

Bicycles constitute an indispensable part of urban transportation and leisure in New York, providing considerable information on the movement of people in the city. The trajectory data that we used for travel flow prediction were from the for-profit public bicycle-sharing system Citi Bike, the largest such system in the USA. We collected 2.8 million bicycle-sharing records in April 2023. The format of the data is shown in Table 1. We divided the region into 1 km × 1 km squares and screened the top 40% of the historical flow, which amounted to 118 squares. The zonal distribution of travel flow is illustrated in Figure 3, which shows significant bicycle traffic in Midtown Manhattan (over 60,000 rides per day). The destinations were largely concentrated in Manhattan and Brooklyn. We divided the bicycle traffic data into weekday and weekend groups for separate analyses.

4.1.2. POI data

We used a Python API to obtain POI data and analyse the impact of human activities on spatial interactions. As shown in Table 2, we divided the POI data into eight major activity categories, such as entertainment, catering, education, and services. We labelled these activity categories

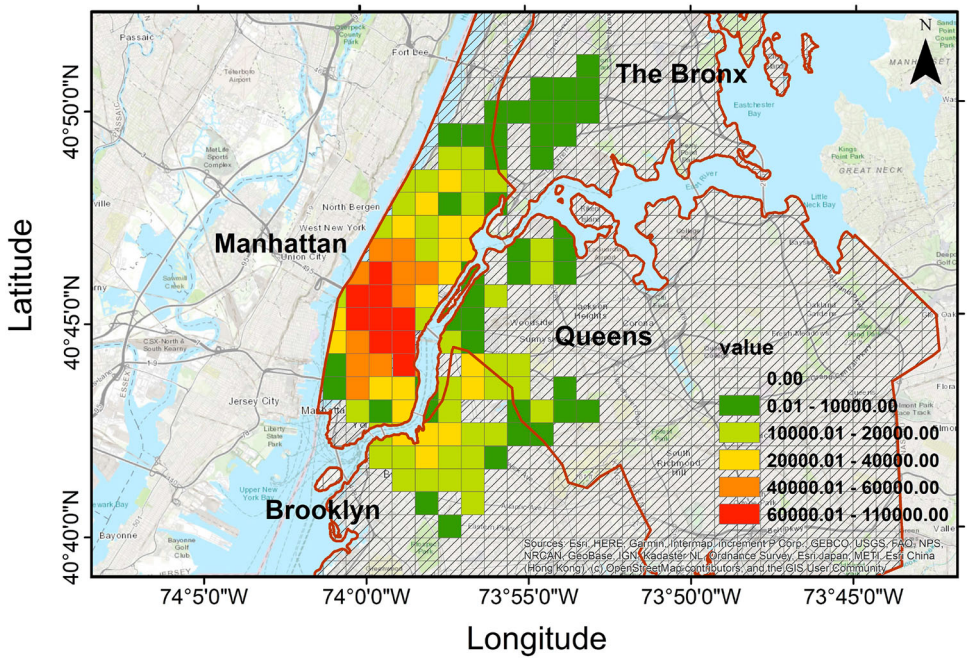

Figure 3. The distribution of bike sharing travel flow in the research area, where the value indicates average daily grid outflows.

Table 2. The description of activities in New York.

Label	Activities	Description
T.01	Entertainment	Leisure and recreational places, such as theatres and resorts.
T.02	Catering	Locations serving food and beverages, such as restaurants, snack bars and cafes, etc.
T.03	Healthcare	Places providing medical and health services, such as nursing homes and hospitals.
T.04	Utilities	Welfare facilities for general communities, such as public parks, sports centres, etc.
T.05	Service	Locations that provide daily services, such as post offices, logistics companies.
T.06	Retail	Shopping or consumption places, such as supermarkets and department stores.
T.07	Education	Locations for cultural dissemination and education training, such as universities.
T.08	Others	Other kinds of human activities or facilities.

T.01, *T*.02, and so on, and calculated the number of each category in predefined squares to reflect the activities in a given area. We integrated these two forms of data, with the final format shown in Table 3.

4.2. Result and analysis

We compared the performance of our model with that of several baseline methods (GRU, T-GCN, GPR, and SPGCL). The prediction results are shown in Table 4. A GRU is a commonly used neural network that can extract traffic patterns from time series. T-GCN is a hybrid deep learning model that successively combines A GCN (to extract spatial correlations) and GRU (for time series

Table 3. The format of input data, where the sample point displays its label, longitude, latitude, timestamp, the outflow of departing bikes at a given time, and the number of POI points in 8 categories inside the grid.

					POI number			
Grid ID	Latitude	Longitude	Timestamp	Outflow	T.01	T.02	···	T.08
D.08	−73.98	40.74	1681346457	366	32	44	···	56

Table 4. The results of the framework with different matrices of New York City.

Time	Method	a	b	β	RMSE	MAE	Accuracy	R^2
Weekdays	GRU	\	\	\	17.79	11.42	71.26	0.88
	T-GCN$_A$	\	\	−0.4	15.57	9.97	74.85	0.90
	GPR-TGCN$_{S'}$	\	\	\	15.79	10.31	75.10	0.92
	GPR-TGCN$_S$	\	\	\	13.60	8.94	78.03	0.93
	GPR-TGCN$_{A,S}$	0.3	0.7	−0.4	14.02	9.13	77.35	0.92
	SPGCL	\	\	\	14.51	9.38	76.55	0.92
	SG-GCN	**0.3**	**0.7**	**−0.4**	**11.99**	**7.85**	**80.62**	**0.94**
Weekends	GRU	\	\	\	17.89	10.96	75.28	0.90
	T-GCN$_A$	\	\	−0.4	16.95	10.75	76.56	0.91
	GPR-TGCN$_{S'}$	\	\	\	17.00	11.33	76.50	0.91
	GPR-TGCN$_S$	\	\	\	15.29	10.13	78.86	0.92
	GPR-TGCN$_{A,S}$	0.5	0.5	-0.4	16.49	10.86	77.20	0.91
	SPGCL	\	\	\	16.76	11.77	76.83	0.91
	SG-GCN	**0.5**	**0.5**	**−0.4**	**14.98**	**9.81**	**79.28**	**0.93**

forecasting). T-GCN$_A$ denotes the prediction accuracy of inputting the distance matrix A as the adjacency matrix into T-GCN. GPR is a non-parametric regression analysis that involves the prior probability distributions of points from Gaussian processes. GPR-TGCN$_{S'}$ and GPR-TGCN$_S$ represent the training results of inputting the matrices S' and S (computed by GPR), respectively, into the T-GCN process. In GPR-TGCN$_{A,S}$, the adjacent matrix is calculated by GPR through a linear combination and then fed into T-GCN. SPGCL (see Section 3) performs self-paced graph contrastive learning simply using the distance matrix A. The proposed SG-GCN framework employs a GCN, a GRU, and GPR, which is specially threshold-filtered through SPGCL based on GPR-TGCN$_{A,S}$. The optimal decay coefficient β of the distance matrix was estimated to be −0.4 on both weekdays and weekends.

We compared the updated kernel integrated with POIs with our original model (see Section 3.3.2) using the same hyperparameters: a learning rate of 0.001, a batch size of 32, 3000 training epochs, and 100 hidden layers. The best-fitted p values of the adjacent matrices were 0.1 on weekdays and 0.05 on weekends. For comparison, the hyperparameters for GCN and GRU in the baseline model were a maximum of 3000 epochs, a learning rate of 0.001, a weight decay of 0, a batch size of 32, and 100 hidden layers.

In the Table, it shows that SG-GCN had the highest accuracy among the tested methods, whereas the GRU method had the lowest accuracy. This suggests that the performance of SG-GCN is superior to that of traditional GCNs, indicating its suitability for wide use in forecasting. For weekdays, GPR-TGCN$_{A,S}$ had the best performance when $a = 0.3$ and $b = 0.7$. Since a and b are the coefficients of the distance and similarity matrices, respectively, this suggests that on weekdays, destinations' characteristics (such as travel volume and activities) have a greater impact on travel flow prediction, while it is difficult to describe spatial interactions based on travel distances. It can therefore be inferred that residents pay more attention to destination characteristics than to travel distances. One explanation is that most residents have weekday work schedules. On weekends, on the other hand, the parameters were best fitted when $a = 0.5$ and $b = 0.5$. This suggests that residents care more about travel distances than about destination characteristics. Since most people do not work on weekends, cycling is more restricted by distance than other modes of transport. These results are reasonable and within our expectations.

The matrices generated by the model are shown in Figure 4, The horizontal and vertical coordinates correspond to the labels of the two sites, and the degree of correlation is indicated by colour value, with darker colours indicating places with stronger correlations. To further investigate the spatial interactions of the areas on travel volumes, we also generated an OD matrix for New York. We took the studied square sequence as the matrix's horizontal and vertical coordinates and then aggregated the OD data by counting the motions between each start–end pair. The entries of the matrix represent the travel volume from square $D.i$ to square $D.j$ and vice versa.

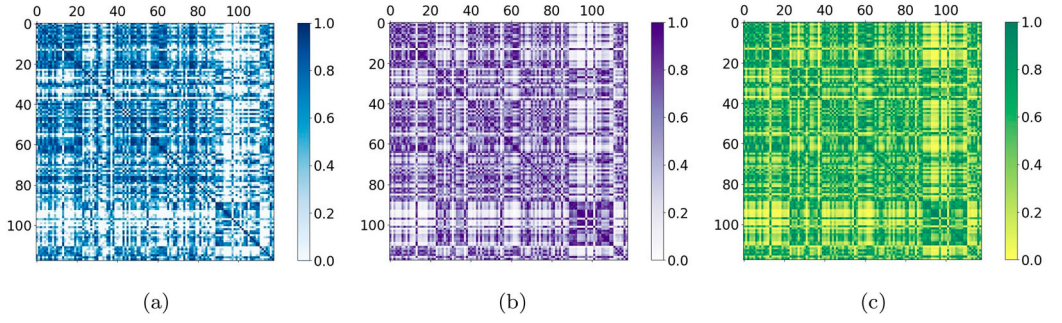

Figure 4. The matrics distribution in New York. (a) OD matrix (b) *A* (c) *S*.

The areas are labelled by activity category. The main activities were education in D.01–D.19, catering in D.20–D.23, healthcare in D.24–D.39, entertainment in D.40–D.52, services in D.53–D.60, retail in D.61–D.79, and utilities in D.80–D.118.

Figure 4 shows that education (D.01–D.19), entertainment (D.40–D.52), and retail areas (D.61–D.79) had strong spatial interactions, with large OD travel volumes (see Figure 4(a)). Since bicycling is strongly constrained by distance, this indicates that residents are likely to cycle through universities and business districts. It can also be inferred that activities such as education and retail have strong convergence relationships. Figure 4(b) shows that D.85–D.110 (utilities) exhibited strong correlations, suggesting that utilities such as parks and sport centres are aggregately located in New York. Figure 4(c) shows that D.85–D.110 (utilities) had strong correlations with most areas in D.24–D.39 (healthcare) in *S*. This is an interesting finding, since these two activities did not exhibit strong correlations in terms of OD bicycle flows (see Figure 4(a)). One explanation is that people who exercise in parks or gyms are likely to sustain injuries and be taken to a hospital directly. However, bicycling is a distance-constrained mode of transportation. Moreover, taking a patient to the hospital by bicycle could be problematic. Therefore, we hypothesise that strong interactions exist between hospitals and recreation areas, which may be difficult to discern through bicycle–sharing data.

5. Case study in Ningbo, China: Taxi trips prediction

5.1. Study area and data

To further validate the proposed framework's effectiveness, we conducted a case study of Ningbo, China. Ningbo is a chemical industry centre in eastern China and one of Zhejiang Province's commercial centres. As in the case of New York, the Ningbo dataset consisted of travel flow and POI data. We collected 1.25 million taxi OD records from March 2017 in a similar format (see Table 1). We selected data from the top 47% of historical flows, totalling 100 squares, to make predictions. Figure 5 shows that the taxi flow distribution in Ningbo was mainly concentrated in the Beilun and especially Haishu and Yinzhou districts. As in the case of New York, we collected POI data for each square divided into 14 major categories. The best-fitted p values of the adjacent matrices in Ningbo were 0.05 on weekdays and 0.1 on weekends. (Table 5)

5.2. Result and analysis

The number of hidden layers in the Ningbo experiment was 150, while the other hyperparameters were the same as in the case of New York. As shown in Table 6, we compared the results of the six methods. SG-GCN had the best performance for both weekdays and weekends, while T-GCN$_A$

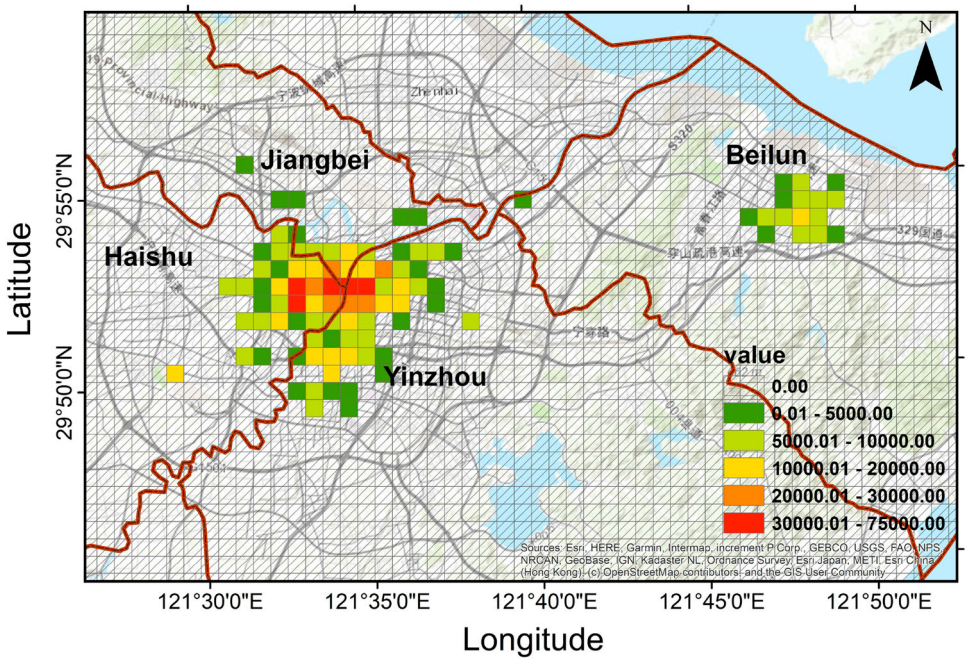

Figure 5. The distribution of taxi travel flow in Ningbo.

Table 5. The description of activities in Ningbo.

Label	Activities	Description
B.01	Entertainment	Leisure and recreational places, such as theatres and resorts.
B.02	Catering	Locations serving food and beverages, such as restaurants, snack bars and cafes, etc.
B.03	Healthcare	Places providing medical and health services, such as nursing homes and hospitals
B.04	Utilities	Locations that provide daily services, such as post offices, logistics companies.
B.05	Banking	Economic and financial institutions, such as banks and ATMs.
B.06	Retail	Shopping or consumption places, such as supermarkets and department stores.
B.07	Education	Locations for cultural dissemination and education training, such as universities.
B.08	Attractions	Scenic sports, such as parks, aquariums.
B.09	Hotels	Places that provide accomodation, such as hotels and lodges.
B.10	Fitness	Sports and fitness places, such as gyms and sports centres.
B.11	Transports	Transportation facilities, such as airports and subway stations.
B.12	Residence	Residential areas, such as apartments and residential complexes.
B.13	Enterprises	All kinds of companies or industrial parks.
B.14	Automobile	Places that provide automobile services, such as car sales centres and repair shops.

showed the worst performance. These results demonstrate the superiority of SG-GCN in both bicycle and taxi travel predictions.

The optimal attenuation coefficient β for the distance matrix was estimated to be -2 on weekdays and -1 on weekends. GPR-TGCN$_{A,S}$ showed the best performance when $a = 0.2$ and $b = 0.8$ on weekdays and when $a = 0.3$ and $b = 0.7$ on weekends.

On weekends, the Ningbo coefficient b was higher than a compared to New York, indicating that location characteristics (activities and travel flows) had a greater influence on passengers' travel behaviours than travel distances did. One explanation is that taxis have considerably fewer distance constraints than bicycles, and passengers can travel farther to more attractive areas. In fact, the distance constraints on weekends were even fewer than on weekdays. It can be inferred that passengers in Ningbo had more time on weekends, and trip distances could be slightly longer. Conversely,

Table 6. The results of the framework with different matrices of Ningbo.

Time	Method	*a*	*b*	*β*	RMSE	MAE	Accuracy	R^2
Weekdays	GRU	\	\	\	7.42	4.73	73.02	0.88
	T-GCN$_A$	\	\	−2	6.95	4.94	74.71	0.89
	GPR-TGCN$_{S'}$	\	\	\	6.84	4.58	75.12	0.90
	GPR-TGCN$_S$	\	\	\	6.31	4.36	77.06	0.91
	GPR-TGCN$_{A,S}$	0.2	0.8	−2	6.49	4.46	76.40	0.91
	SPGCL	\	\	\	6.54	4.63	76.22	0.91
	SG-GCN	**0.2**	**0.8**	−2	**6.07**	**4.23**	**77.92**	**0.92**
Weekends	GRU	\	\	\	8.80	5.31	75.07	0.90
	T-GCN$_A$	\	\	−1	8.59	5.72	75.65	0.90
	GPR-TGCN$_{S'}$	\	\	\	8.37	5.84	76.27	0.91
	GPR-TGCN$_S$	\	\	\	7.90	5.65	77.59	**0.91**
	GPR-TGCN$_{A,S}$	0.3	0.7	−1	7.91	5.47	77.56	0.91
	SPGCL	\	\	\	7.26	4.88	79.40	0.93
	SG-GCN	**0.3**	**0.7**	**-1**	**6.52**	**4.49**	**81.51**	**0.94**

most residents in New York covered much shorter distances on weekends than on weekdays (the *a* values were 0.5 for weekends and 0.3 for weekdays). Since most retail locations are closed in New York on weekends, it can be inferred that these travel preferences are influenced by culture.

As in the case of New York, we calculated the effects of the three types of matrices in Ningbo with the same hyperparameters: a learning rate of 0.001, a batch size of 32, and 150 hidden layers. To establish functional areas in order to simplify the analysis, we also classified the squares by the number of POI categories present in each square. The main activity categories were catering in D.01–D.08, transport in D.09–D.10, utilities in D.11–D.19, retail in D.20–D.91, enterprises in D.92–D.99, and automobile services in D.100.

Figure 6 shows that D.20–D.32 and D.33–D.44 exhibited the strongest interactions. Almost all values were above 0.7, indicating a larger taxi flow within these two functional zones. The main activities in both areas were related to shopping. The figure shows that the interaction between these two areas was stronger than in New York, which can be attributed to the overlapping activities of the two areas, resulting in a stronger preference for shopping in business and trading areas. Furthermore, a comparison between weekdays and weekends shows that the spatio-temporal correlations of retail zones were more strongly influenced by time differences than in New York. Also, the correlation between shops was significantly stronger on weekends than on weekdays, which can be attributed to the more pronounced demarcation between work and leisure time in Ningbo, which makes people more inclined to go shopping by taxi on weekends. Conversely, shops and public utilities in New York generally close earlier on weekends, resulting in lower similarity between weekdays and weekends.

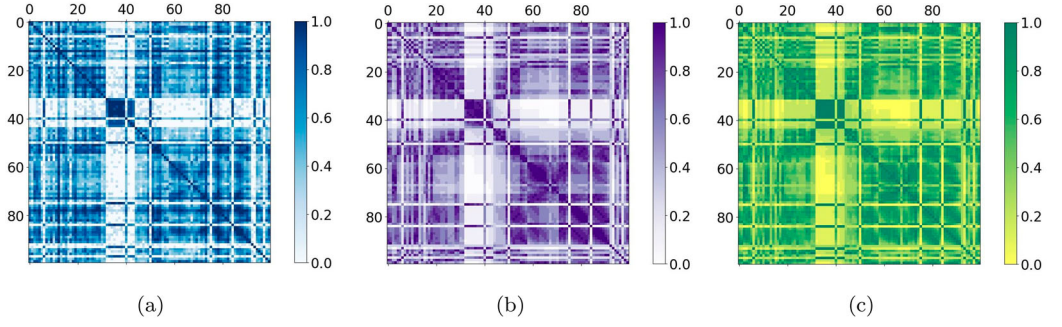

Figure 6. The matrics distribution in Ningbo. (a) OD matrix (b) *A* (c) *S*.

6. Conclusion

In this study, we innovatively developed a Gaussian-based GCN framework and SPGCL to predict travel flows. Specifically, we used a non-stationary Gibbs kernel parameterised with an input-dependent length scale to construct spatial relationship matrices that incorporate the temporal dimension. Moreover, SPGCL automatically determines the spatial interaction threshold, which was previously unattainable by GCN models performing traffic flow prediction tasks. The proposed framework showed better performance in predicting urban travel flows than baseline models. To assess the framework's performance, we conducted two case studies of New York City, USA, and Ningbo, China, using 2.8 million bicycle-sharing records and 1.25 million taxi trip records, respectively. The results show that our framework makes more accurate predictions and can be used in a wide range of human mobility prediction tasks.

The results also reveal that the main activity types and distances between areas influence the patterns of spatial interconnections. Entertainment and shopping activities are more plentiful and have a strong impact on travel behaviours. Movements between locations with the same types of activities is generally lower than between areas with different types of activities. Furthermore, areas for exercise in New York (such as parks and fitness centres) exhibit strong interactions with hospitals, while their bicycle flow interactions are very weak. This may be attributed to the fact that intense physical activity may result in injuries, necessitating a visit to a medical facility. However, the transition from a fitness centre to a medical facility typically does not involve the use of bicycles. This is an interesting finding that could not be obtained using other methods.

One limitation of our work is that we focussed on technique improvements for travel volume forecasting using GPS data and did not consider other types of data, such as street view images. In the future, we aim to test the performance of our framework using multiple types of data. Our findings have far-reaching implications for investigating spatial interactions and identifying patterns of crowd motion. Managers can reveal the semantics of spatial interactions under various circumstances by accurately predicting inter-area travel flows over time, thereby optimising intelligent transportation and urban planning. Moreover, our framework may be adapted to various city sizes and modes of transportation, which is important for appropriately managing people's preferences for activity locations.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is supported by the Fundamental Research Funds for the Central Universities [Grant No.2-9-2022-060], the Beijing Natural Science Foundation [Grant No. 8222009].

ORCID

Rui Cao http://orcid.org/0000-0002-1440-4175
Changfeng Jing http://orcid.org/0000-0002-1270-5353
Yu Liu http://orcid.org/0000-0002-0016-2902

References

Ahmed, M. S., and A. R. Cook. 1979. "Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques, No. 722." https://api.semanticscholar.org/CorpusID:106553179.

Chen, X., Z. Wang, Q. Hua, W. L. Shang, Q. Luo, and K. Yu. 2022. "Ai-Empowered Speed Extraction Via Port-Like Videos for Vehicular Trajectory Analysis." *IEEE Transactions on Intelligent Transportation Systems* 24:4541–4552. https://doi.org/10.1109/TITS.2022.3167650.

Chen, X., S. Wu, C. Shi, Y. Huang, Y. Yang, R. Ke, and J. Zhao. 2020. "Sensing Data Supported Traffic Flow Prediction Via Denoising Schemes and Ann: A Comparison." *IEEE Sensors Journal*20:14317–14328. https://doi.org/10.1109/JSEN.7361.

Cheng, Z., J. Caverlee, K. Lee, and D. Sui. 2011. "Exploring Millions of Footprints in Location Sharing Services." In *Proceedings of the International AAAI Conference on Web and Social Media*, 81–88. Buffalo, NY: The AAAI Press.

Chi, G., J. C. Thill, D. Tong, L. Shi, and Y. Liu. 2016. "Uncovering Regional Characteristics From Mobile Phone Data: A Network Science Approach." *Papers in Regional Science* 95 (3): 613–631. https://doi.org/10.1111/pirs.12149.

Eysenbach, G., and J. Wyatt. 2002. "Using the Internet for Surveys and Health Research." *Journal of Medical Internet Research* 4:e862.

Feng, X., X. Ling, H. Zheng, Z. Chen, and Y. Xu. 2018. "Adaptive Multi-Kernel Svm with Spatial–temporal Correlation for Short-Term Traffic Flow Prediction." *IEEE Transactions on Intelligent Transportation Systems* 20:2001–2013. https://doi.org/10.1109/TITS.6979.

Fotheringham, A. S. 1981. "Spatial Structure and Distance-Decay Parameters." *Annals of the Association of American Geographers* 71 (3): 425–436. https://doi.org/10.1111/anna.1981.71.issue-3.

Gong, S., J. Qin, H. Xu, R. Cao, Y. Liu, C. Jing, Y. Hao, and Y. Yang. 2023. "Spatio-Temporal Parking Occupancy Forecasting Integrating Parking Sensing Records and Street-Level Images." *International Journal of Applied Earth Observation and Geoinformation* 118:103290. https://doi.org/10.1016/j.jag.2023.103290.

Gong, S., Z. Sun, Y. Zuo, and T. Bian. 2022. "Spatio-Temporal Travel Volume Prediction and Spatial Dependencies Discovery Using GRU, GCN and Bayesian Probabilities. In *2022 7th International Conference on Big Data Analytics (ICBDA)*, 130–136. IEEE.

Gonzalez, M. C., C. A. Hidalgo, and A. L. Barabasi. 2008. "Understanding Individual Human Mobility Patterns." *Nature* 453:779–782. https://doi.org/10.1038/nature06958.

Gray, C., and V. Mueller. 2012. "Drought and Population Mobility in Rural Ethiopia." *World Development* 40 (1): 134–145. https://doi.org/10.1016/j.worlddev.2011.05.023.

Guo, D., X. Zhu, H. Jin, P. Gao, and C. Andris. 2012. "Discovering Spatial Patterns in Origin-Destination Mobility Data." *Transactions in GIS* 16 (3): 411–429. https://doi.org/10.1111/tgis.2012.16.issue-3.

Huang, X., Y. Zhao, S. Wang, X. Li, D. Yang, Y. Feng, Y. Xu, L. Zhu, and B. Chen. 2022. "Unfolding Community Homophily in Us Metropolitans Via Human Mobility." *Cities (London, England)*129:103929.

Koesdwiady, A., R. Soua, and F. Karray. 2016. "Improving Traffic Flow Prediction with Weather Information in Connected Cars: A Deep Learning Approach." *IEEE Transactions on Vehicular Technology* 65:9508–9517. https://doi.org/10.1109/TVT.2016.2585575.

Kong, X., Q. Chen, M. Hou, H. Wang, and F. Xia. 2023. "Mobility Trajectory Generation: A Survey." *Artificial Intelligence Review* 56 (S3): 3057–3098. https://doi.org/10.1007/s10462-023-10598-x.

Kong, X., K. Wang, M. Hou, F. Xia, G. Karmakar, and J. Li. 2022. "Exploring Human Mobility for Multi-Pattern Passenger Prediction: A Graph Learning Framework." *IEEE Transactions on Intelligent Transportation Systems* 23:16148–16160. https://doi.org/10.1109/TITS.2022.3148116.

Kraemer, M. U., C. H. Yang, B. Gutierrez, C. H. Wu, B. Klein, D. M. Pigott, and O. C. D. W. Group, et al. 2020. "The Effect of Human Mobility and Control Measures on the COVID-19 Epidemic in China." *Science (New York, N.Y.)* 368 (6490): 493–497. https://doi.org/10.1126/science.abb4218.

Lalchand, V., K. Tazi, T. M. Cheema, R. E. Turner, and S. Hosking. 2022. "Kernel Learning for Explainable Climate Science." Preprint arXiv:2209.04947.

Li, R., T. Zhong, X. Jiang, G. Trajcevski, J. Wu, and F. Zhou. 2022. "Mining Spatio-Temporal Relations Via Self-Paced Graph Contrastive Learning." In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 936–944. New York: Association for Computing Machinery.

Lin, L., Z. He, and S. Peeta. 2018. "Predicting Station-Level Hourly Demand in a Large-Scale Bike-Sharing Network: A Graph Convolutional Neural Network Approach." *Transportation Research Part C: Emerging Technologies* 97:258–276. https://doi.org/10.1016/j.trc.2018.10.011.

Lin, G., A. Lin, and D. Gu. 2022. "Using Support Vector Regression and K-nearest Neighbors for Short-Term Traffic Flow Prediction Based on Maximal Information Coefficient." *Information Sciences*608:517–531. https://doi.org/10.1016/j.ins.2022.06.090.

Liu, K., S. Gao, and F. Lu. 2019. "Identifying Spatial Interaction Patterns of Vehicle Movements on Urban Road Networks by Topic Modelling." *Computers, Environment and Urban Systems* 74:50–61. https://doi.org/10.1016/j.compenvurbsys.2018.12.001.

Liu, Y., X. Liu, S. Gao, L. Gong, C. Kang, Y. Zhi, G. Chi, and L. Shi. 2015. "Social Sensing: A New Approach to Understanding Our Socioeconomic Environments." *Annals of the Association of American Geographers* 105 (3): 512–530. https://doi.org/10.1080/00045608.2015.1018773.

Liu, Y., Z. Sui, C. Kang, and Y. Gao. 2014. "Uncovering Patterns of Inter-Urban Trip and Spatial Interaction From Social Media Check-In Data." *PloS One* 9:e86026. https://doi.org/10.1371/journal.pone.0086026.

Medina-Salgado, B., E. Sánchez-DelaCruz, P. Pozos-Parra, and J. E. Sierra. 2022. "Urban Traffic Flow Prediction Techniques: A Review." *Sustainable Computing: Informatics and Systems* 35:100739.

Miller, H. J. 2004. "Tobler's First Law and Spatial Analysis." *Annals of the Association of American Geographers* 94 (2): 284–289. https://doi.org/10.1111/j.1467-8306.2004.09402005.x.

Rasmussen, C. E. 2004. "Gaussian Processes in Machine Learning." In *Advanced Lectures on Machine Learning: ML Summer Schools 2003*, Revised Lectures, 63–71. Canberra, Australia, Tübingen, Germany: Springer.

Rasmussen, C. E., and C. K. Williams. 2006. *Gaussian Processes for Machine Learning*. Vol. 1. Cambridge, Massachusetts: Springer.

Roy, J. R., and J. C. Thill. 2003. "Spatial Interaction Modelling." *Papers in Regional Science* 83:339–361. https://doi.org/10.1007/s10110-003-0189-4.

Russell, S. J. 2010. *Artificial Intelligence a Modern Approach*. London: Pearson Education, Inc.

Shan, Z., D. Zhao, and Y. Xia. 2013. "Urban Road Traffic Speed Estimation for Missing Probe Vehicle Data Based on Multiple Linear Regression Model." In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, 118–123. The Hague, Netherlands: IEEE.

Sîrbu, A., G. Andrienko, N. Andrienko, C. Boldrini, M. Conti, F. Giannotti, R. Guidotti, et al. 2021. "Human Migration: The Big Data Perspective." *International Journal of Data Science and Analytics*11 (4): 341–360. https://doi.org/10.1007/s41060-020-00213-5.

Smith, B. L., and M. J. Demetsky. 1994. "Short-Term Traffic Flow Prediction Models-A Comparison of Neural Network and Nonparametric Regression Approaches." In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, 1706–1709. San Antonio, TX: IEEE.

Sutskever, I., O. Vinyals, and Q. V. Le. 2014. "Sequence to Sequence Learning with Neural Networks." In *Advances in Neural Information Processing Systems*, Vol. 27. Montreal, Quebec: Curran Associates, Inc.

Tu, W., R. Cao, Y. Yue, B. Zhou, Q. Li, and Q. Li. 2018. "Spatial Variations in Urban Public Ridership Derived From Gps Trajectories and Smart Card Data." *Journal of Transport Geography* 69:45–57. https://doi.org/10.1016/j.jtrangeo.2018.04.013.

Wang, Y., J. Zheng, Y. Du, C. Huang, and P. Li. 2022. "Traffic-Ggnn: Predicting Traffic Flow Via Attentional Spatial-Temporal Gated Graph Neural Networks." *IEEE Transactions on Intelligent Transportation Systems* 23:18423–18432. https://doi.org/10.1109/TITS.2022.3168590.

Williams, B. M., and L. A. Hoel. 2003. "Modeling and Forecasting Vehicular Traffic Flow As a Seasonal Arima Process: Theoretical Basis and Empirical Results." *Journal of Transportation Engineering* 129 (6): 664–672. https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(664).

Williams, C., and C. Rasmussen. 1995. "Gaussian Processes for Regression." In *Advances in Neural Information Processing Systems*, Vol. 8. Cambridge, MA: MIT Press.

Zhang, Q., C. Huang, L. Xia, Z. Wang, Z. Li, and S. Yiu. 2023. "Automated Spatio-Temporal Graph Contrastive Learning." In *Proceedings of the ACM Web Conference 2023*, 295–305. Austin, TX: Association for Computing Machinery.

Zhang, L., Q. Liu, W. Yang, N. Wei, and D. Dong. 2013. "An Improved K-nearest Neighbor Model for Short-Term Traffic Flow Prediction." *Procedia-Social and Behavioral Sciences* 96:653–662. https://doi.org/10.1016/j.sbspro.2013.08.076.

Zhang, J., Y. Zheng, and D. Qi. 2017. "Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction." In *Proceedings of the AAAI Conference on Artificial Intelligence*. San Francisco, California: AAAI Press.

Zhao, L., Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li. 2020. "T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction." *IEEE Transactions on Intelligent Transportation Systems* 21:3848–3858. https://doi.org/10.1109/TITS.6979.

Zheng, W., D. H. Lee, and Q. Shi. 2006. "Short-Term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach." *Journal of Transportation Engineering* 132 (2): 114–121. https://doi.org/10.1061/(ASCE)0733-947X(2006)132:2(114).

Zhou, T., B. Huang, X. Liu, G. He, Q. Gou, Z. Huang, and C. Xie. 2020. "Spatiotemporal Exploration of Chinese Spring Festival Population Flow Patterns and Their Determinants Based on Spatial Interaction Model." *ISPRS International Journal of Geo-Information* 9:670. https://doi.org/10.3390/ijgi9110670.