Modeling bike-sharing demand using a regression model with spatially varying coefficients

Xudong Wang [a,c], Zhanhong Cheng [a,c], Martin Trépanier [b,c], Lijun Sun [a,c,*]

[a] *Department of Civil Engineering, McGill University, Montreal, QC H3A 0C3, Canada*
[b] *Department of Mathematics and Industrial Engineering, Polytechnique Montreal, Montreal, QC H3T 1J4, Canada*
[c] *Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT), Montreal, QC H3T 1J4, Canada*

ARTICLE INFO

ABSTRACT

As an emerging mobility service, bike-sharing has become increasingly popular around the world. A critical question in planning and designing bike-sharing services is to know how different factors, such as land-use and built environment, affect bike-sharing demand. Most research investigated this problem from a holistic view using regression models, where assume the factor coefficients are spatially homogeneous. However, ignoring the local spatial effects of different factors is not tally with facts. Therefore, we develop a regression model with spatially varying coefficients to investigate how land use, social-demographic, and transportation infrastructure affect the bike-sharing demand at different stations to address this problem. Unlike existing geographically weighted models, we define station-specific regression and use a graph structure to encourage nearby stations to have similar coefficients. Using the bike-sharing data from the BIXI service in Montreal, we showcase the spatially varying patterns in the regression coefficients and highlight more sensitive areas to the marginal change of a specific factor. The proposed model also exhibits superior out-of-sample prediction power compared with traditional machine learning models and geostatistical models.

1. Introduction

Growing concerns about urban sustainability and climate change have led to increasing interest in green transportation solutions such as bike-sharing (Shaheen et al., 2010). Bike-sharing systems can reduce air pollution and natural resource consumption (Cai et al., 2019), improve public health (Fishman et al., 2013), and support multimodal transport connections by acting as a "last mile" connection to public transport (DeMaio, 2009). Because of these advantages, many cities have established bike-sharing systems. By 2014, the number of cities that operate bike-sharing programs is 855, with a total of 946,000 bikes in operation (Fishman, 2016), and the numbers have increased to over 1500 cities and 4.5 million bikes recently (Fishman and Allan, 2019).

Most modern bike-sharing systems are station-based in which users can borrow and return bicycles at specified docking stations (DeMaio, 2009), such as the BIXI system in Montreal, Canada. Lots of studies have shown factors, such as weather, built environment and land-use, public transportation, socio-demographic attributes, and temporal factors, play an essential role in bike-sharing demand (Eren and Uz, 2020).

Understanding how different factors affect user demand at a station level is essential to the planning and operation of a bike-sharing service, because planners/operators can make different strategies according to the specific factors. More importantly, comprehending the correlations between these factors and user demand can help us estimate the potential demand for new stations in advance to assistant service planning.

Much research has studied the relationship between station-level bike-sharing demand and various factors by regression models. Early studies used a simple global regression model (e.g., multivariate linear regression) with fixed coefficients for all stations (Rixey, 2013; Faghih-Imani et al., 2014; El-Assi et al., 2017). This approach neither considers spatial correlation nor captures the heterogeneous effects of factors to different stations. Therefore, spatial autoregression (SAR) models (Zhang et al., 2017; Shen et al., 2018; Guidon et al., 2020) and spatially varying coefficients (SVC) models (Bao et al., 2018; Munira and Sener, 2020; Yang et al., 2020) are applied to address the limitations of the simple global regression. Both SAR and SVC models utilize spatial dependencies, while only SVC models can explain the heterogeneous effects of a specific factor on different stations.

Inspired by the network-lasso problem (Hallac et al., 2015), we propose a new SVC model to investigate how influential factors contribute to the bike-sharing demand spatially. We impose a graph regularization to a set of station-specific linear regression models to enable adjacent stations sharing similar coefficients. This model's fundamental is entirely new to traditional SVC models, such as the geographically weighted regression (GWR) (Brunsdon et al., 1998) and Gaussian-process-based models (Gelfand et al., 2003; Banerjee et al., 2008; Lindgren et al., 2011). Specifically, our approach has a more precise form and can scale to large problems. Besides, the graph regularization is more flexible to problems involving non-Euclidean distance (Wu et al., 2020).

We apply the graph regularized SVC method to bike-sharing demand from BIXI, a bike-sharing system in Montreal, Canada. The dependent variable is the average hourly departure demand at each bicycle station. The bike-sharing demand exhibits dissimilarity in temporal, so we analyze the demand of the morning peak and afternoon peak separately. We select twelve factors related to demographics, land-use, transportation infrastructures, and bicycle facilities as candidate independent variables. Then, we apply Pearson correlation analysis on these candidate factors to choose the most relevant variables. Unlike only using circle buffers to aggregate the factors in most studies, we combine Thiessen polygons (Edelsbrunner and Seidel, 1986) and circle buffers instead to obtain non-overlapping catchment areas of each station.

We evaluate the performance of proposed SVC model from both regression and prediction aspects. (1) In terms of regression, we find coefficients' spatial distributions given by the proposed SVC model are consistent with real-life observations. Moreover, the proposed SVC model greatly reduces the regression residual and the spatial autocorrelation in the residual, compared with a global linear regression. (2) In terms of prediction, we apply the proposed SVC model to predict the potential demand for new (unobserved) stations. Compared with the traditional machine learning models and geostatistical models, the proposed model shows superior out-of-sample prediction power.

We summarize the main contributions of this work as follows:

- We introduce a graph regularization to station-specific linear regressions, working as a new SCV model.
- We analyze the effect of land-use, social-demographic, transportation infrastructure, and bicycle facilities on bike-sharing demand at a station-level by the proposed SCV model.
- We combine circle buffers and Thiessen polygons to extract factors from non-overlapping catchment areas, which improves the bike-sharing demand modeling.
- We demonstrate the proposed SVC model's applicability in interpretability and predicting the potential demand for new stations.

The remainder of this paper is organized as follows. In Section 2, we review related work on bike-sharing demand modeling. Section 3 introduces the data of bicycle demand and influential factors. Section 4 presents the regression model with spatially varying coefficients, where we use a graph regularization to encourage nearby stations to have similar coefficients. Section 5 demonstrates the regression result and compares the proposed method with other models. Section 6 concludes this study and discusses some directions for future research.

2. Related work

Over the past decades, numerous research has analyzed the relationship between bicycle demand and factors such as weather, land-use, and social demographics. We review a particular branch of them, where a regression model was used to analyze the *station-level demand* in *bike-sharing* systems. A comprehensive review can be found in Eren and Uz (2020).

Buck and Buehler (2012) inspected the effect of bike facilities, demographics, and land-use factors on the average daily bike-sharing

check-outs in Washington D.C. by a linear regression model. Rixey (2013) compared the effects of similar factors to the average monthly bike-sharing demand in three US cities. Faghih-Imani et al. (2014) analyzed hourly bike-sharing demand in Montreal with additional consideration of weather and temporal factors. A linear mixed model was used to capture the dependencies between repeated observations of the same station; a similar technique was also applied to a Toronto bike-sharing system (El-Assi et al., 2017). More studies are based on a similar method to evaluate the effect of weather, temporal factors, building environment, and social demographics on bike-sharing demand in different cities (Miranda-Moreno and Nosal, 2011; Noland et al., 2016; Mateo-Babiano et al., 2016; Scott and Ciuro, 2019).

The spatial dependency for bike-sharing demand is not sufficiently utilized in the research mentioned above. Traditionally, there are two main approaches to incorporate spatial dependency: 1) adding spatially lagged dependent variable as additional covariates; 2) assuming a spatial auto-correlation process in the regression residuals (e.g., regression kriging). Here we collectively refer to these two as spatial autoregressive (SAR) models. Many studies have applied SAR models to model bike-sharing demand. For example, Zhang et al. (2017) studied the spatial correlations in bike-sharing demand and quantified the spatial correlation in demand between nearby stations. They found a significant spatial correlation in the bike-sharing demand. Faghih-Imani and Eluru (2016) further incorporated the spatially and temporal interactions into the bike-sharing demand modeling. Shen et al. (2018) explored the factors affecting the usage of dockless bikes in nine consecutive days by a SAR model. The spatial impact was introduced by a spatial weight matrix defined as the inverse distance among neighborhoods within 5 km. Similarly, Guidon et al. (2020) compared a SAR model with random forests in bike-sharing demand modeling; results showed that SAR outperformed random forests because of the ability to incorporate spatial dependence.

Although SAR models improve the regression performance by spatial dependence, the regression coefficients are still the same for all stations. Therefore, SAR is also inadequate in depicting the heterogeneous effect of factors on different stations. To address this problem, a few studies applied spatially varying coefficients (SVC) models in bike-sharing demand modeling. For instance, Bao et al. (2018) used a classical SVC model–geographically weighted regression (GWR)–to capture the spatial heterogeneity of the effect of points of interest (POI) to bike-sharing demand. GWR is a weighted local regression model where the distance between two observations determines the weights. Based on the similar idea, Munira and Sener (2020) analyzed how socioeconomic and land-use influence Strava bike activity by geographically weighted Poisson regression (GWPR). All coefficients in traditional GWR are local variables; the semi-parametric GWR (S-GWR) model can further incorporate both local and global variables in the regression model. In this direction, Yang et al. (2019, 2020) used S-GWR to explore the spatially varying relationship in Chicago. From the literature, few studies used another SVC model apart from GWR. This paper thus analyzes the bike-sharing demand by a new SVC framework based on a graph-regularization.

3. Data

This section introduces the data applied in the research. Section 3.1 presents the bicycle demand data obtained from BIXI, a Montreal bike-sharing system. The bike-sharing related factors are described in Section 3.2, we will evaluate the effect of these factors on bike-sharing demand.

3.1. Bike-sharing demand

BIXI, the first large-scale bike-sharing system in North America, is located in Montreal, Canada. In 2019, there are 615 stations and 5.6 million trip records in the BIXI system. Because of the cold winter in
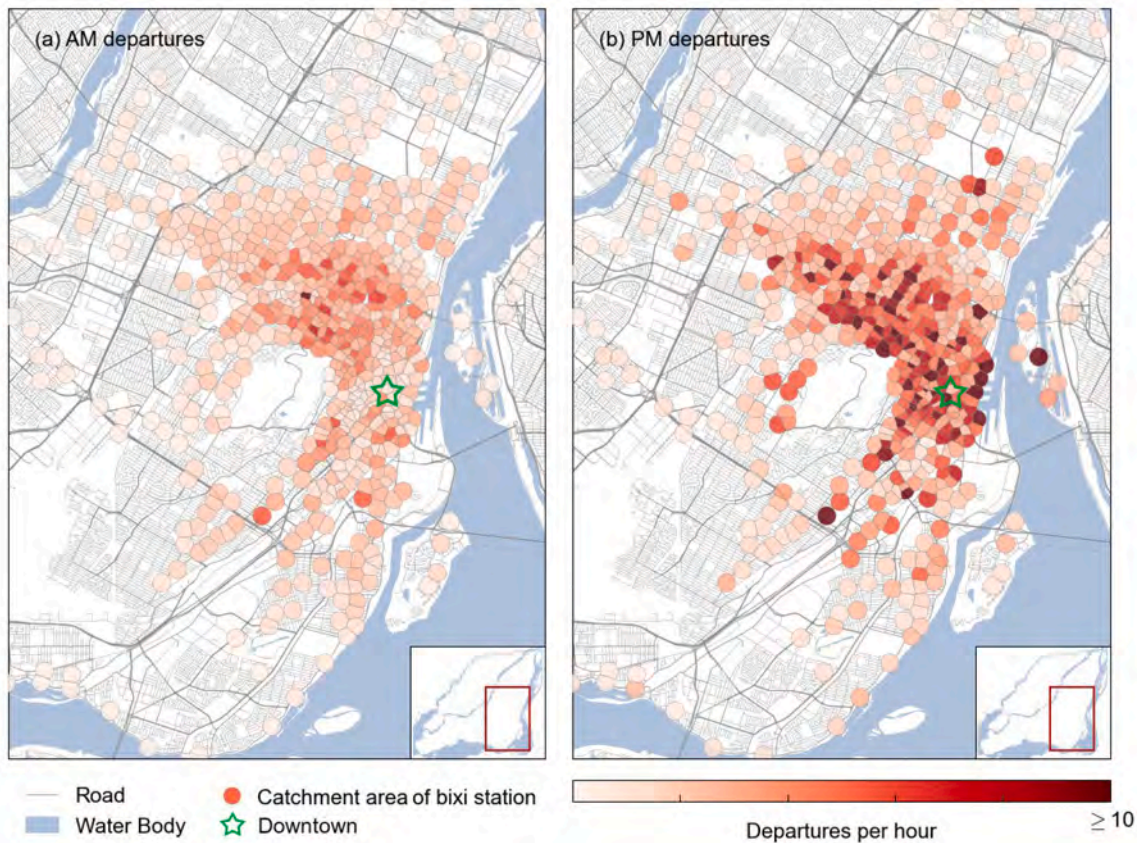
Fig. 1. Average hourly departures for BIXI stations in Montreal (left: morning peak, right: afternoon peak).

Montreal, BIXI only operates from April to November each year. We use the trip data between May and October of 2019 from BIXI to obtain the bicycle ridership since stations are often under adjustments in the beginning and end months of this system (Faghih-Imani et al., 2014). The historical data recorded six attributes for each bicycle trip: origin and destination stations, departure and arrival time, trip duration, and membership information. This data set is publicly available from BIXI (https://bixi.com).

For most stations, bicycle demand peaks at two periods on a weekday—6:00–10:00 am and 3:00–7:00 pm. The spatial distributions of bicycle demand are different in these two periods (Fig. 1). Therefore, we build two models to quantify the effect of bicycle demand factors in the morning and afternoon peaks, respectively. Fig. 1 shows the average hourly departure demand distributions for morning and afternoon peaks, where each BIXI station is represented by a catchment area with 250 m radius, and Thiessen polygon is used to determine the boundary between overlapping catchment areas (refer to Section 3.2 for more details). In the morning, the northwest side of downtown (Le Plateau-Mont-Royal) has high departure demand; this is a residential area with high population density. In the afternoon, the center of bicycle demand shifts toward the downtown area. Several stations along the river (such as the Old Port of Montreal, a popular sightseeing area) and near Rosemont area with friendly ride paths have significantly high demand. In general, afternoon's hourly departure is higher than that in the morning, which conforms with the findings by Faghih-Imani et al. (2014). Another interesting observation is that the distribution of bicycle demand is not smooth in the space; there are many demand hot spots that either stand-alone or gather along a street. This phenomenon could be caused by factors such as metro stations, bicycle paths, and commercial streets around the bike stations.

Table 1
Summary of candidate independent variables[a] (aggregated by catchment areas).

Factors	Description	Min	Max	Mean
Population	Total residential population.	0	4042	1073.88
Commercial	The number of commercial POI.	0	89	14.94
Service	The number of services POI.	0	55	12.43
Government	The number of government POI.	0	13	1.17
Park	The ratio of parks in the catchment area.	0	1	0.08
University	A binary variable for university.	0	1	0.01
Metro	A binary variable for metro station.	0	1	0.13
Bus	log(number of bus routes + 1).	0	3.3	1.21
Walkscore	A measure of walkability from 0 to 100.	7	99	79.28
Road lenth	Total length of roads (in meters).	133.27	4824.40	1934.93
Cycle path	The proportion of roads with cycle path.	0	1	0.23
Capacity	The number of docks of a BIXI station.	11	105	23.10

[a] Finally selected variables are shown in bold print.

3.2. Influential factors

There are numerous factors, including land-usage, transportation infrastructure, social demographics, and weather conditions, that impact bicycle usage (Rixey, 2013; Faghih-Imani et al., 2014; Zhang et al., 2017; El-Assi et al., 2017). As in this paper, we focus on general demand level at stations; the weather and temporal factors – which affect short-term demand fluctuation – are not considered in this study. Other factors form independent variables in our regression model.

Most previous studies aggregated influential factors by catchment areas with a certain radius around bicycle stations. However, the

Table 2
Pearson correlations between normalized factors and demand.

	Pop	Comm	Service	Gov	Park	Univ	Metro	Bus	Walk	Road	Cycle	Capacity
Pop	1.0**											
Comm	0.14**	1.0**										
Service	0.4**	0.66**	1.0**									
Gov	0.14**	0.21**	0.23**	1.0**								
Park	−0.14**	−0.25**	−0.21**	−0.06	1.0**							
Univ	−0.01	0.03	0.02	0.04	−0.03	1.0**						
Metro	−0.09**	0.13**	0.09**	0.03	−0.05	0.08**	1.0**					
Bus	0.13**	0.25**	0.29**	0.18**	−0.03	0.09**	0.34**	1.0**				
Walk	−0.11**	0.35**	0.24**	0.16**	−0.32**	0.06	0.05	0.03	1.0**			
Road	0.51**	0.02	0.27**	0.02	0.01	−0.05	0.09**	0.28**	−0.43**	1.0**		
Cycle	−0.1**	−0.15**	−0.16**	−0.13**	0.18**	0.02	−0.07*	−0.14**	−0.1**	−0.21**	1.0**	
Capacity	−0.27**	−0.04	−0.2**	0.03	0.15**	0.14**	0.07*	0.06	0.05	−0.2**	0.07	1.0**
AM demand	−0.01	0.14**	0.12**	−0.01	−0.02	0.1**	0.04	0.07*	0.48**	−0.34**	0.16**	0.11**
PM demand	−0.35**	0.23**	−0.00	0.01	0.07*	0.15**	0.26**	0.04	0.46**	−0.47**	0.17**	0.44**

* Significant at 0.1 level.
** Significant at 0.05 level.

catchment areas can overlap if bicycle stations are close to each other (Fig. 1), which leads to correlations between independent variables in nearby stations. In practice, users tend to choose the bicycle station that is nearest to their origin/destination regardless of the distance of other walkable stations. Therefore, we use the intersection of 250 m circles and Thiessen polygons (Edelsbrunner and Seidel, 1986) to obtain non-overlapping catchment areas, as shown in Fig. 1. The radius of 250 m has been proven to be suitable for bike-sharing demand modeling (Faghih-Imani et al., 2014). Thiessen polygons guarantee every point within the catchment area of a station i is closest to the station i compared with all other stations. Section 5.4 shows that applying Thiessen polygons improves the regression performance compared with only using circle buffer.

We then aggregate candidate independent variables at each catchment area, as summarized in Table 1.These candidate variables will be further selected based on the Pearson correlation analysis in Table 2. The population is estimated from the demographic data obtained from 2016 Canada census data at a dissemination block level. Land-use factors include university, park, and various types of points of interest (POI). POI data are obtained from DMTI Enhanced Point of Interest (DMTI Spatial Inc., 2019). Based on standard industrial classification (SIC, 2020), we obtain three types of POI: commercial (Division G and H), service (Division I), and governments (Division J). There are few other (heavy) industries since the study area is around the city. Factors related to transportation infrastructures include road length, walk score (Walk Score, 2020), metro station, and bus route, where walk score is an index scale from 0 to 100 measuring the walkability to neighboring amenities. Note that we use the logarithm of the bus route number because of its long tail feature. Finally, we use the cycle path and station capacity as two bicycle-related factors. The numbers in Table 1 are in original unit. Due to the giant difference in magnitude, we normalize all variables and user demand to 0 to 1 before applying them to the regression model.

We analyze Pearson correlations of normalized data to filter insignificant factors, as shown in Table 2. Commercial POI, walk sore, cycle path, and station capacity are four factors that show the strongest positive correlation to the bike-sharing demand. On the other hand, Road length shows the most negative correlation to the bike-sharing demand. We remove the government-type POI since it is not significant to both the AM and the PM bike-sharing demand at 0.1 level. Besides, the service-type POI shows a very high correlation (0.66) to the commercial-type POI. Therefore, the service-type POI is also removed to prevent multicollinearity. The finally adopted variables are those shown in bold print in Table 1.

4. Model

4.1. Regression with graph regularization

Let N be the total number of bicycle stations. For station i ($i \in \{1, …, N\}$), y_i denotes its bike-sharing demand, which is the dependent variable. Denote a vector of independent variables by $\mathbf{x}_i = [1, x_{i1}, …, x_{im}]^T$ for station i, where m is the number of factors (as shown in Table 1). The regression model with spatially varying coefficients is described as:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta}_i + \varepsilon_i, \quad i = 1, …, N, \tag{1}$$

where $\boldsymbol{\beta}_i = [\beta_{i0}, \beta_{i1}, …, \beta_{im}]^T$ is a coefficient vector for station i, ε_i is the error term. We can estimate the coefficients for all the stations by minimizing the sum of squared errors:

$$min \sum_{i=1}^{N} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_i \right)^2. \tag{2}$$

If coefficient vectors $\boldsymbol{\beta}_i$ are the same for all stations, Eq. (1) becomes a simple linear regression model and function (2) is the least square problem. However, we study the factors' effect at a station-level and assume coefficient vectors are varying over stations. In this situation, Eq. (1) is not identifiable since the number of unknown coefficients $N \times (m + 1)$ are much larger than the number of equations N; one can find many sets of $\boldsymbol{\beta}_i$ that minimizes function (2) (to zero). One solution for addressing the problem is that we can impose some constraints on $\boldsymbol{\beta}_i$ to make the regression model solvable.

A fundamental assumption in modeling the spatial effects is that nearby stations have similar coefficients. This assumption is based on the first law of geography (Tobler, 1970): "everything is related to everything else, but near things are more related than distant things." Many SAR and SVC models are based on this assumption. Following this assumption, we introduce a graph structure similar to the *network lasso* problem (Hallac et al., 2015). Consider the bike-sharing stations on a graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$, where $\mathscr{V}$ and $\mathscr{E}$ are the set of nodes and edges, respectively. Each node represents a bicycle station ($|V| = N$), and an edge represents a connection between two nodes. Instead of a fully connected graph, we assume each station is connected to its K nearest stations with undirected edges, where K controls the number of adjacent stations. By doing so, the specific local information of each station is incorporated into the model, and the computation time can also be substantially reduced. There are other possible approaches to build a graph, such as using distance-based methods or a Gabriel graph (Gabriel and Sokal, 1969). This paper applied the k-nearest neighbor graph as an example.

The linear regression model for each station with a graph regularization (GR) term, which penalizes the difference between $\boldsymbol{\beta}_i$ in adjacent nodes, is proposed in:

$$min \sum_{i=1}^{N} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_i \right)^2 + \lambda \sum_{(i,j) \in \mathscr{E}} w_{ij} \left\| \boldsymbol{\beta}_i - \boldsymbol{\beta}_j \right\|_2^2, \quad \lambda \geq 0, \tag{3}$$

where the parameter λ balances the regression error and the difference between coefficients in adjacent nodes, w_{ij} is the weight of the edge (i,j), which is a decaying function of distance. Here we apply $w_{ij}(\alpha) = d_{ij}^{-\alpha}$ with $\alpha > 0$, where d_{ij} is the distance between node i and j, and α controls the decaying speed. When $\lambda = 0$, each node has its own optimization without considering other nodes. Increasing λ encourages the neighboring nodes to have similar coefficients. With a sufficiently large λ, function (3) degrades to one (or several) linear regression model(s) with all stations in each connected component[1] of the graph sharing the same coefficients. Note we assume each independent variable has a station-specific coefficient in (3). However, this framework can be easily extended to a hierarchical structure with both fixed and varying coefficients, like the semi-parametric GWR model (Yang et al., 2020), by adjusting the graph regularization term.

Function (3) is a convex optimization problem and can be efficiently solved using optimization software, such as CVXPY. For large-scale problems, the alternating direction method of multipliers (ADMM) can be used to solve the problem in a distributed and scalable manner (Hallac et al., 2015).

For a new station p with unknown bike-sharing demand, we can estimate its factor coefficients $\boldsymbol{\beta}_p$ by interpolating the coefficients $\boldsymbol{\beta}^\star$ from known stations and then achieve demand prediction of station p. Let Nei(p) denote the K nearest stations of station p, and we want to find a $\boldsymbol{\beta}_p$ that minimizes the difference between the coefficients of neighbors, leading to the following optimization:

$$min \sum_{q \in \text{Nei}(p)} w_{pq} \left\| \boldsymbol{\beta}_p - \boldsymbol{\beta}_q^\star \right\|_2^2. \tag{4}$$

Function (4) can also be efficiently solved by convex optimization, and the prediction ability of a model can work as an indicator to measure the performance of the model.

In summary, the advantages of the proposed GR model are mainly

twofold: (i) make function (2) solvable by adding a graph regularization term into the linear regression model and obtain spatially varying coefficients as a consequence to investigate the factors' effect on demand at station-level; (ii) have a flexible structure to be adjusted for different problems.

4.2. Hyper-parameter tuning

The regression with graph regularization model has three hyper-parameters: λ determines the intensity of the graph regularization term, K is the number of neighbors of a node, and α controls the weight decaying of edges. We use 10-fold cross-validation to search for the optimal values of hyper-parameters. First, the whole data set (demand and corresponding factors) is normalized by min-max feature scaling to bring all values into the range $[0, 1]$. Next, all stations are randomly partitioned into 10 equal-sized groups. In each iteration of the cross-validation, retain one group as the validation set $\mathscr{V}_{\text{validate}}$, the remaining stations belong to the training set $\mathscr{V}_{\text{train}}$. For a given combination of hyper-parameters, the coefficients of training stations are estimated by (3), and the coefficients of test stations are obtained by (4). Then, the root mean square error (RMSE) is calculated by the demand prediction on the validation set $\mathscr{V}_{\text{validate}}$:

$$\text{RMSE} = \sqrt{\frac{\sum_{i \in \mathscr{V}_{\text{validate}}} \left(y_i - \widehat{y}_i \right)^2}{|\mathscr{V}_{\text{validate}}|}}, \tag{5}$$

where $\widehat{y}_i = \mathbf{x}_i^T \boldsymbol{\beta}_i^\star$ is the predicted bike-sharing demand for station i. Repeat the cross-validation process for 10 times and the performance of a hyper-parameter setting is evaluated by the average RMSE of the 10-fold cross-validation. The general procedure of 10-fold cross-validation of searching optimal regularization parameters is summarized in Algorithm 1.

Algorithm 1. 10-fold cross-validation for searching optimal hyper-parameters.

Algorithm 1 10-fold cross-validation for searching optimal hyper-parameters

---

Input: Normalized factor data $X \in \mathbb{R}^{N \times (m+1)}$, normalized bike-sharing demand $\mathbf{y} \in \mathbb{R}^N$, regularization parameters set λ_{set}, link weight parameter set α_{set}, number of neighbors set K_{set}.
Output: Optimal regularization parameters $\lambda^\star$, optimal link weight parameter $\alpha^\star$, optimal neighbor number $K^\star$
1: Split the data set $\mathcal{D} = \{X, \mathbf{y}\}$ into 10 equal-sized groups;
2: **for** each $\lambda \in \lambda_{set}$, $\alpha \in \alpha_{set}$ and $K \in K_{set}$ **do**
3:     **for** each $j = 1, \ldots, 10$ **do**
4:         $\mathcal{V}_{\text{Validation}}$ = the j th group data of $\mathcal{D}$;
5:         $\mathcal{V}_{\text{Train}}$ = the remaining data of $\mathcal{D}$;
6:         Apply CVXPY on $\mathcal{V}_{\text{Train}}$ in (3) to get $\boldsymbol{\beta}_{\text{train}}^\star$;
7:         Apply CVXPY on $\mathcal{V}_{\text{Validation}}$ in (4) to get $\boldsymbol{\beta}_{\text{validation}}^\star$;
8:         Get prediction $\hat{\mathbf{y}}$ on $\mathcal{V}_{\text{Validation}}$ in Eq. (1);
9:         $\epsilon_j \leftarrow$ RMSE in Eq. (5).
10:     **end for**
11:     $E_{\lambda, \alpha, K} \leftarrow \sum_{j=1}^{10} \epsilon_j / 10$.
12: **end for**
13: $\lambda^\star, \alpha^\star, K^\star \leftarrow \arg\min\{E_{\lambda, \alpha, K}\}$.
14: **return** $\lambda^\star$, $\alpha^\star$ and $K^\star$

---

[1] A connected component of an undirected graph is a subgraph in which any two nodes are connected to each other by paths, and which is not connected to nodes other than this component.

To determine the search scope of hyper-parameters, we first discuss the effect of each hyper-parameter. Fig. 2 illustrates the effect of λ from one round cross-validation, where we use the afternoon peak data and set $K = 4$ and $\alpha = 1$. It can be seen that when λ is close to 10^{-3}, the RMSE
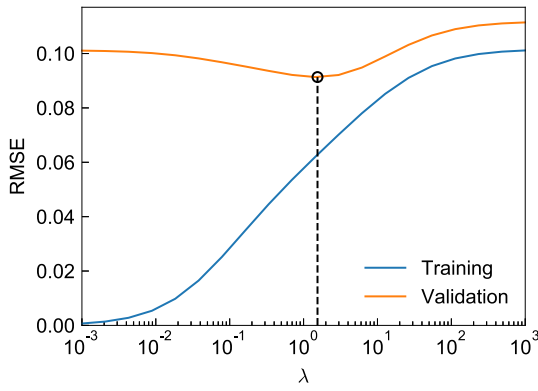
Fig. 2. The effect of λ to the RMSE of training and validation set, using demand of afternoon peak, $K = 4$, $\alpha = 1$.
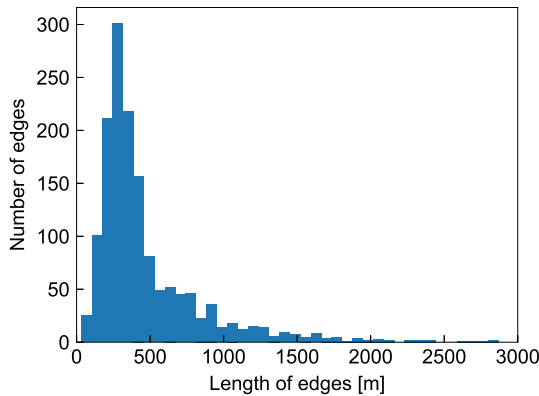

Fig. 3. Histogram of the length of edges.

on the training set is almost 0 because each node is optimized independently. However, the model is over-fitting and generalizes poorly to the validation set with high RMSE. When λ is very large, the impact of graph regularization is much stronger, which means the model reduces to a global regression with invariant coefficients. As a consequence, large λ increases the RMSE for both the training and validation sets. To get the best performance on the validation set, we need to find a λ that appropriately balances the regression error and the difference between neighbors, as marked by the circle with the minimal RMSE on the validation set in Fig. 2.

K determines the number of neighbors of a node. When K is very small (1 or 2), the graph may have too many separate components, which impedes information to be shared among stations and even results

in an ill-posed problem. A too-large K (e.g., $K \approx N$) is also inappropriate since it connects far away (less relevant) stations and also increases the computational burden. Based on the experimental results, a proper interval for K in this problem is between 3 and 7.

Compared with λ and K, α is a less important parameter. The effect of α on the RMSE of the validation set is marginal and can be compensated by tuning a proper λ, because these two hyper-parameters together determine the graph regularization term. We also test exponential-type weight decaying function $w_{ij}(\alpha) = \exp(-\alpha d_{ij})$ with $\alpha > 0$, similar observations are found. This is different from models like GWR, where the bandwidth – controls the speed of weight decaying – plays an important role. A possible explanation could be: the graph regularization is a local constraint, and the edges between a node and its neighbors are of similar distances (small variance); therefore, when α changes, the weights of edges of a node change near proportionally.

Based on the above analysis, we set $\alpha = 1$ and perform a grid search on λ from 0.5 to 10 at 0.5 interval, $K \in \{3,4,5,6,7\}$ by 10-fold cross-validation. We select $\lambda = 2$, $K = 4$ for the morning peak and $\lambda = 2.5$, $K = 4$ for the afternoon peak, when the minimal average RMSE of the 10-fold cross-validation is achieved.

Finally, we investigate the graph when $K = 4$. The number of nodes $|\mathcal{V}| = 615$, and the number of edges $|\mathcal{E}| = 1485$. Note $|\mathcal{E}|/|\mathcal{V}| = 2.4$ is slightly denser than a two-dimensional grid graph. Fig. 3 shows the histogram of edges' length. We can see a clear long-tail distribution. Most connected stations are within 500 m of distance, with a peak at around 250 m, which accords with the 250-m radius of the catchment area. This is because the system is designed to be accessible within a walking distance. Nonetheless, a few stations are far away from other stations. In fact, there are six stations that are disconnected from the other 609 stations. In general, the graph when $K = 4$ achieves a balance between connectivity and redundancy.

5. Results

We apply the proposed graph regularized regression (GR) to the BIXI bike-sharing data and exhibit the result from four aspects. Firstly, using the distribution of coefficients, we analyze how different factors affect stations' demand from a global view. Next, we select three specific stations to explain the advantage of the GR model in detail. Further, we use Moran's I test to diagnose the spatial auto-correlation of regression residuals. Finally, we compare the ability of GR with other models in predicting the demand for new stations.

5.1. Coefficients spatial distribution

The factor coefficients are estimated based on the optimal hyper-parameters from Section 4.2. The numerical distribution for each coefficient in our model is shown as the box plots in Fig. 4, along with a
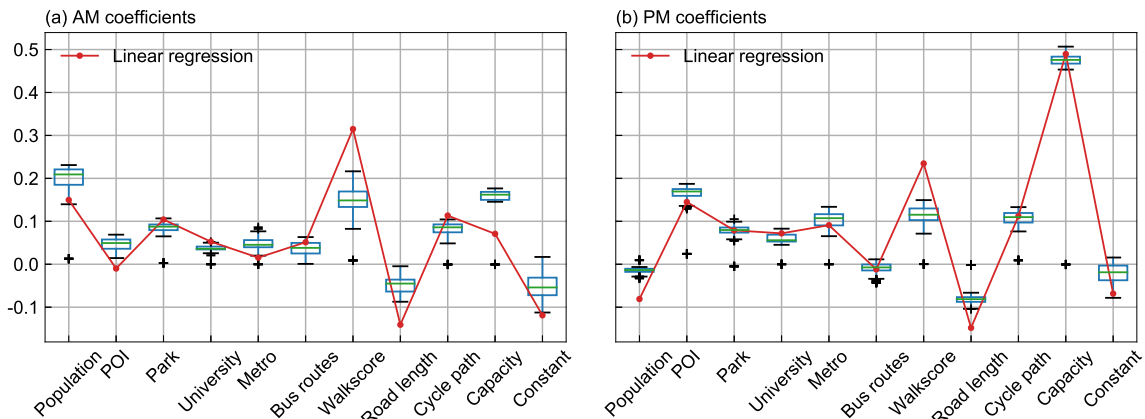

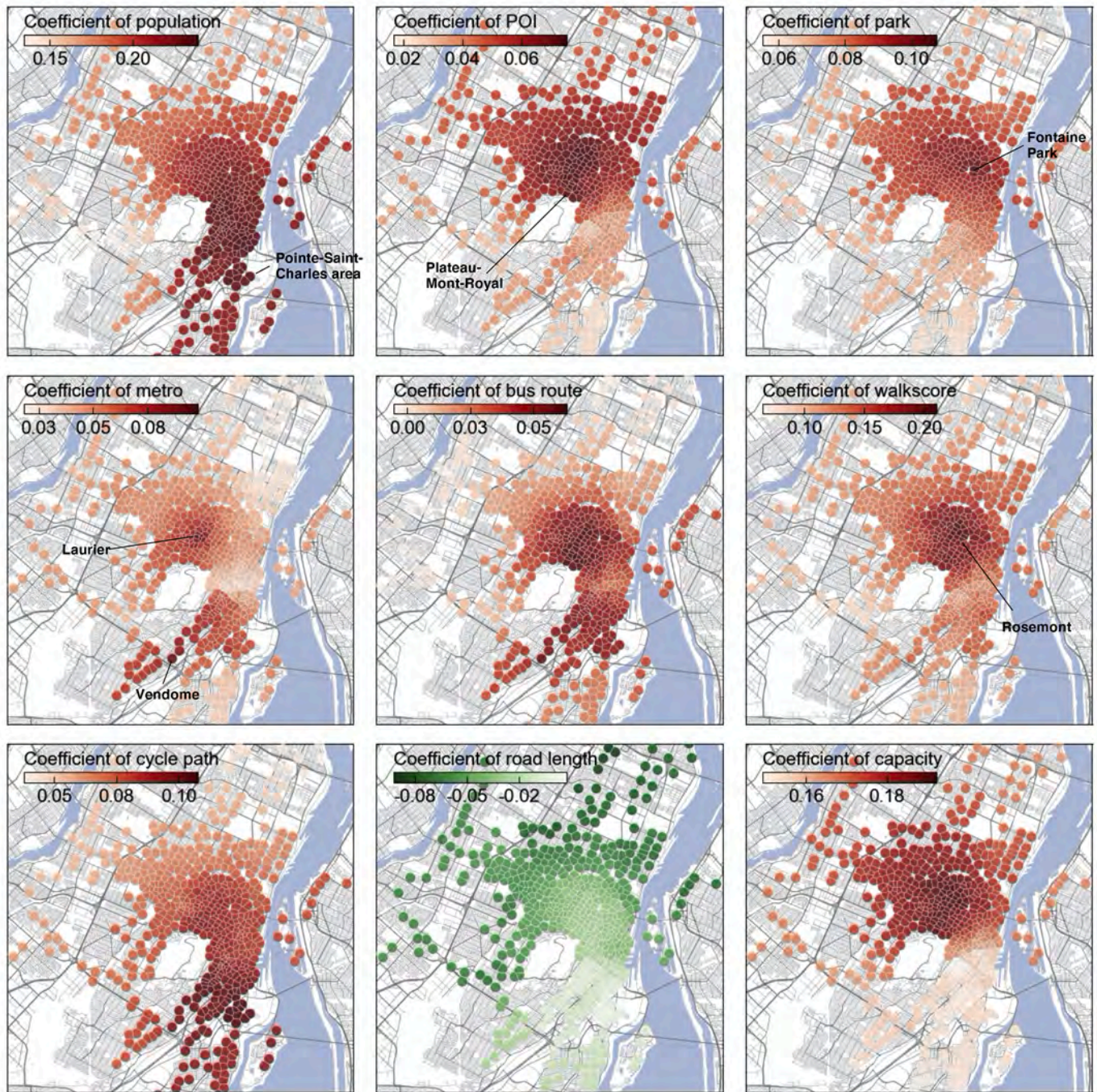**Fig. 4.** The box plot of factor coefficients for morning peak (left) and afternoon peak (right).

Fig. 5. The factor coefficients distribution of morning peak with optimal parameters $\lambda = 2$, $\alpha = 1$ and $K = 4$.

global linear regression (LR) model shown by a red line. The box of each factor depicts the Q1 to Q3 quartiles of the varying coefficients from the proposed graph regularized regression model (GR); the whiskers show the range of these coefficients. It can be seen that the general result of GR (sign and scale of coefficients) is consistent with the linear regression. Actually, linear regression is a particular case of our model when the graph is fully connected and λ is large enough, which constrains all the stations to have the same factor coefficients. Note that some outliers in the coefficient distributions are caused by a small group of disconnected remote stations in the graph, whose bike-sharing demands are very low.

To get more insights from the result, we first clarify the meaning of a coefficient value. Since the data set has been normalized before applying Algorithm 1, a coefficient represents how many changes to the demand when there is a marginal change in the corresponding (normalized)

factor. We refer to this as the importance/effect of this factor to the bike-sharing demand.

In Fig. 4, linear regression shows walkscore has the largest AM co-efficient, while GR model indicates population is the dominating factor for the bike-sharing demand in the morning. This is because a linear regression coefficient is constant and is a compromise of stations with different patterns. In fact, most people depart from their homes in the morning; it is thus reasonable to see the residential population plays a more important role in affecting the morning demand. The station capacity has the most significant effect on the demand in the afternoon. There are two reasons: (1) the demand in the afternoon is much larger than that in the morning (see Fig. 1), a shortage of supply may occur, and the real demand could be constrained by the capacity (Gammelli et al., 2020); (2) the station capacity is designed to satisfy the demand
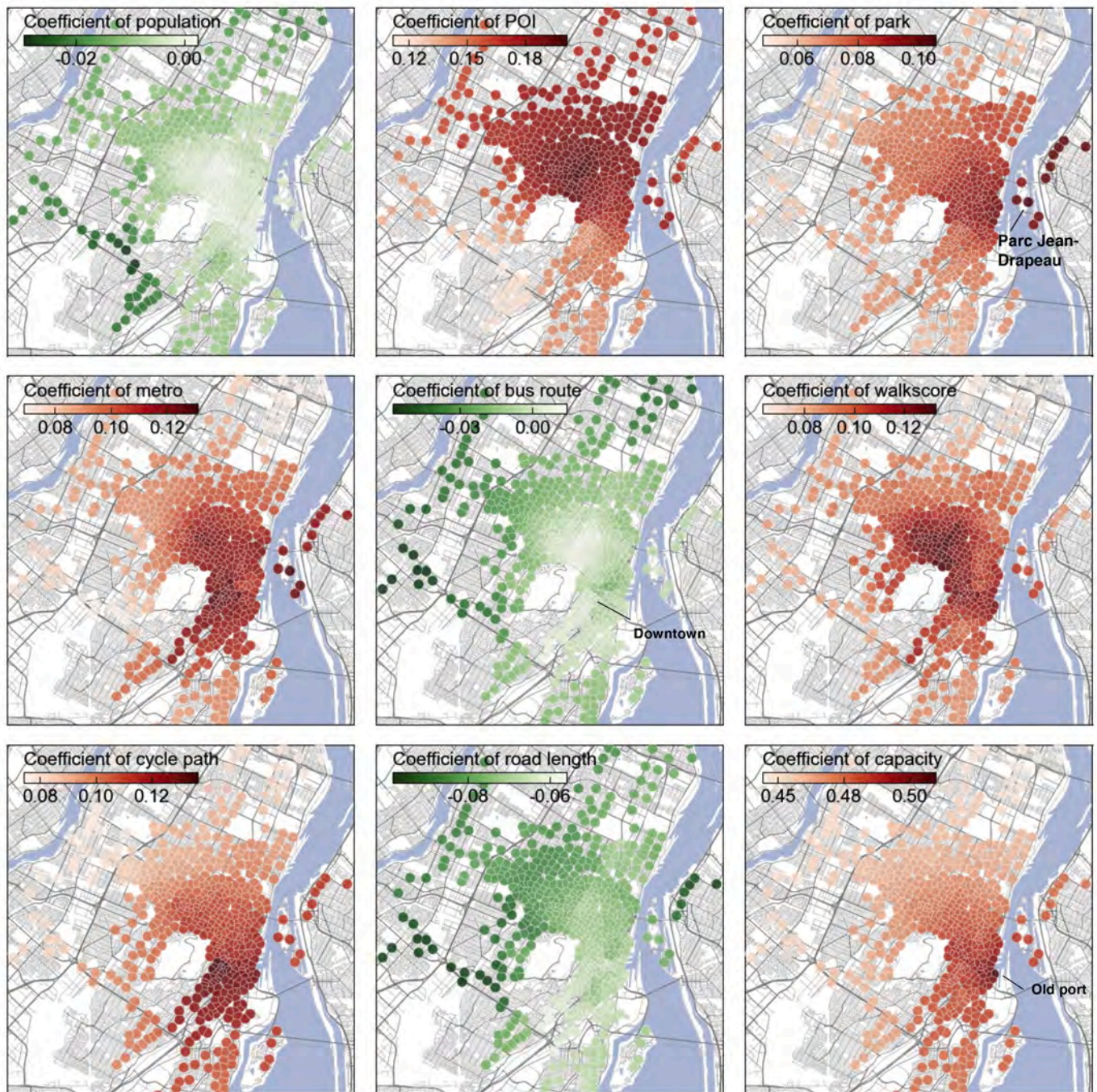
Fig. 6. The factor coefficients distribution of afternoon peak with optimal parameters $\lambda = 2.5$, $\alpha = 1$ and $K = 4$.

pattern, and therefore exhibits a high correlation to the demand in the afternoon peak. Besides population and capacity, the commercial POI coefficient also has an evident difference between the morning and the afternoon. Specifically, the coefficient is larger in the afternoon than that in the morning. It can be explained by more users go for commercial activity (e.g., shopping, entertainment) in late time, which is in accordance with the reality.

We also visualize the spatial distributions for coefficients in the morning and the afternoon peaks by Fig. 5 and Fig. 6, where red markers mean positive values and green markers mean negative values. The coefficient distribution of university is not shown, because there are only 8 (1.3%) stations near a university. Therefore, the university factor has little effect for most stations ($x_{uni} = 0$), but the numerical values for the coefficients of university can be found in Fig. 4.

In the morning, the population coefficients increase toward Pointe-Saint-Charles area and downtown area. Pointe-Saint-Charles, a lively community with many parks, bike paths, and new housing units, has the largest coefficient of population. Also, there are plenty of residential apartments in the downtown area. As for POI, bus route and walk score, we can see the Plateau-Mont-Royal area and a part of Rosemont-La Petite-Patrie area is the coefficient center. These areas are famous for their commercial streets, delightful parks and attractive culture. Specifically, La Fontaine Park, a large compound park, has the highest park coefficient. The cycle path factor (defined as the proportion of roads with cycle path) positively affects the demand, while the road length has a negative effect. It indicates that the automobile-oriented road design is not friendly to bicycles.

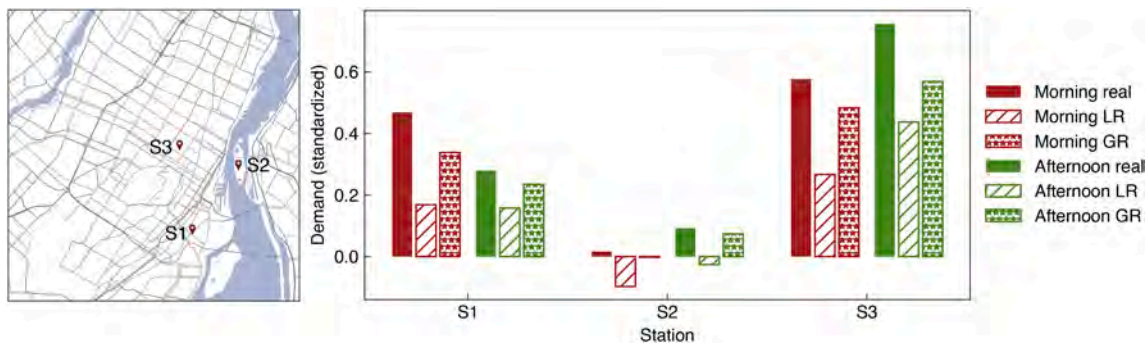On the other side, the factor coefficients can also be used to detect

Fig. 7. The location of selected stations (left) and the estimated demand by LR and GR (right).
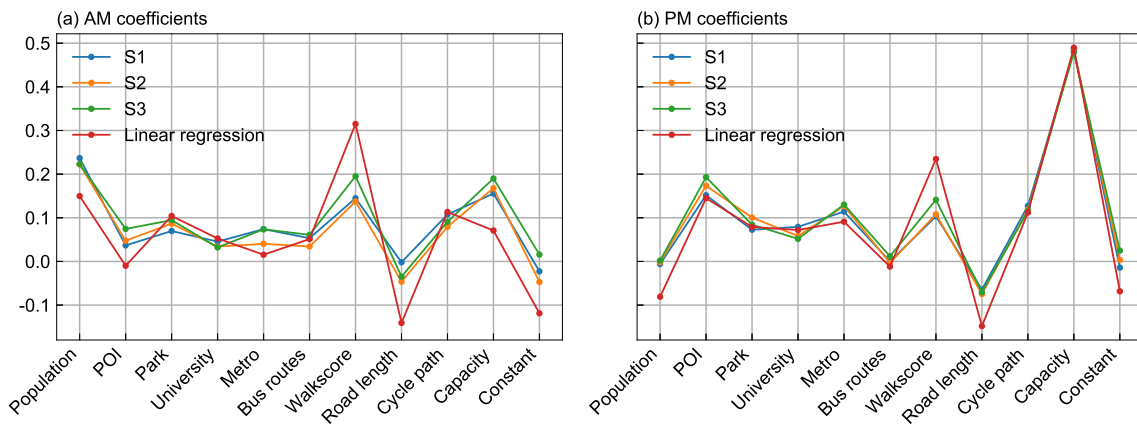

Fig. 8. The factor coefficients of LR and three stations by GR.

"unusual stations". For example, there are two unusual points for the metro coefficients shown in Fig. 5, one is next to the metro station Laurier, another is next to the metro station Vendôme. These two stations have significantly higher metro coefficients than their neighbors. It is because both stations are the transport hub and serve the Orange Line, and the result is consistent with their too high metro coefficients and bus route coefficients.

Fig. 6 shows the spatial distributions for factor coefficients in the afternoon peak. Compared with the coefficients of the morning, except road length, the number of residential population and bus routes also negatively correlates to the departure demand in the afternoon, which means more these two factors would have less bicycle demand. This may indicate an uneven distribution of the residential and work areas in Montreal. In the afternoon, the POI, walk score and cycle path still positively impact the Plateau-Mont-Royal area and a part of Rosemont-La Petite-Patrie area.

The stations near Parc Jean-Drapeau and the Old Port are two special stations in the afternoon. Parc Jean-Drapeau, an isolated island situated to the east of downtown Montreal, has many attractions, including an amusement park, casinos, environmental museum. The metro station Jean-Drapeau is the main entrance to enter the island except by driving. It is reasonable that this station has a high park coefficient and metro coefficient. The Old Port of Montreal is one of the most popular tourist resorts, attracting millions of people. So the capacity of bike stations is quite essential for the demand at this location.

5.2. Examples of estimated demand

We select three stations to further explain how the spatially varying coefficients impact the BIXI demand compared with the global linear regression model. The locations of the selected stations are marked in Fig. 7, where S1 is near Metro Charlevoix in the Pointe-Saint-Charles

area, S2 is on the Parc Jean-Drapeau, and S3 is around Metro Mont-Royal in the Plateau-Mont-Royal area. The three stations represent four kinds of demand: morning peak station (S1) vs. afternoon peak station (S3) and low demand station (S2) vs. high demand station (S3).

The factor coefficients of each station are in Fig. 8. LR model is a unified spatial model, where all the stations have the same factor coefficients (red line in Fig. 8). The proposed spatially varying coefficients GR model takes spatial heterogeneity into account; therefore, each station has its unique factor coefficients (other lines in Fig. 8). From Fig. 7, the LR model shows lower performance on estimating demand in the three stations. Moreover, the LR model can obtain negative estimated demand when the real demand is low (S2). However, the GR model can estimate a better demand in high/low demand or morning/afternoon peak demand.

Although the factor coefficients are slight difference to S1, S2 and S3 in Fig. 8, it matters to estimate the bike-sharing demand (Fig. 7) and reflects the factor effect on demand for each station as discussed in Section 5.1. For example, S1 and S3 have larger metro coefficients comparing with S2 in the morning because S1 and S3 are near the metro while S2 on a sightseeing island.

5.3. Regression residuals

Besides interpreting regression coefficients, it is also important to diagnose the regression residuals. Fig. 9 shows the regression residuals of LR and GR. The residuals of GR are much smaller than LR for both the morning and the afternoon cases. Particularly, the residuals of LR in the morning (upper left of Fig. 9) exhibit strong spatial autocorrelation — negative values for the center of the study area and positive for the peripheral area. In contrast, the spatial autocorrelation is greatly relieved by the GR model. For the afternoon case, the residual patterns of LR and GR are more similar. The outputs of both LR and GR are lower
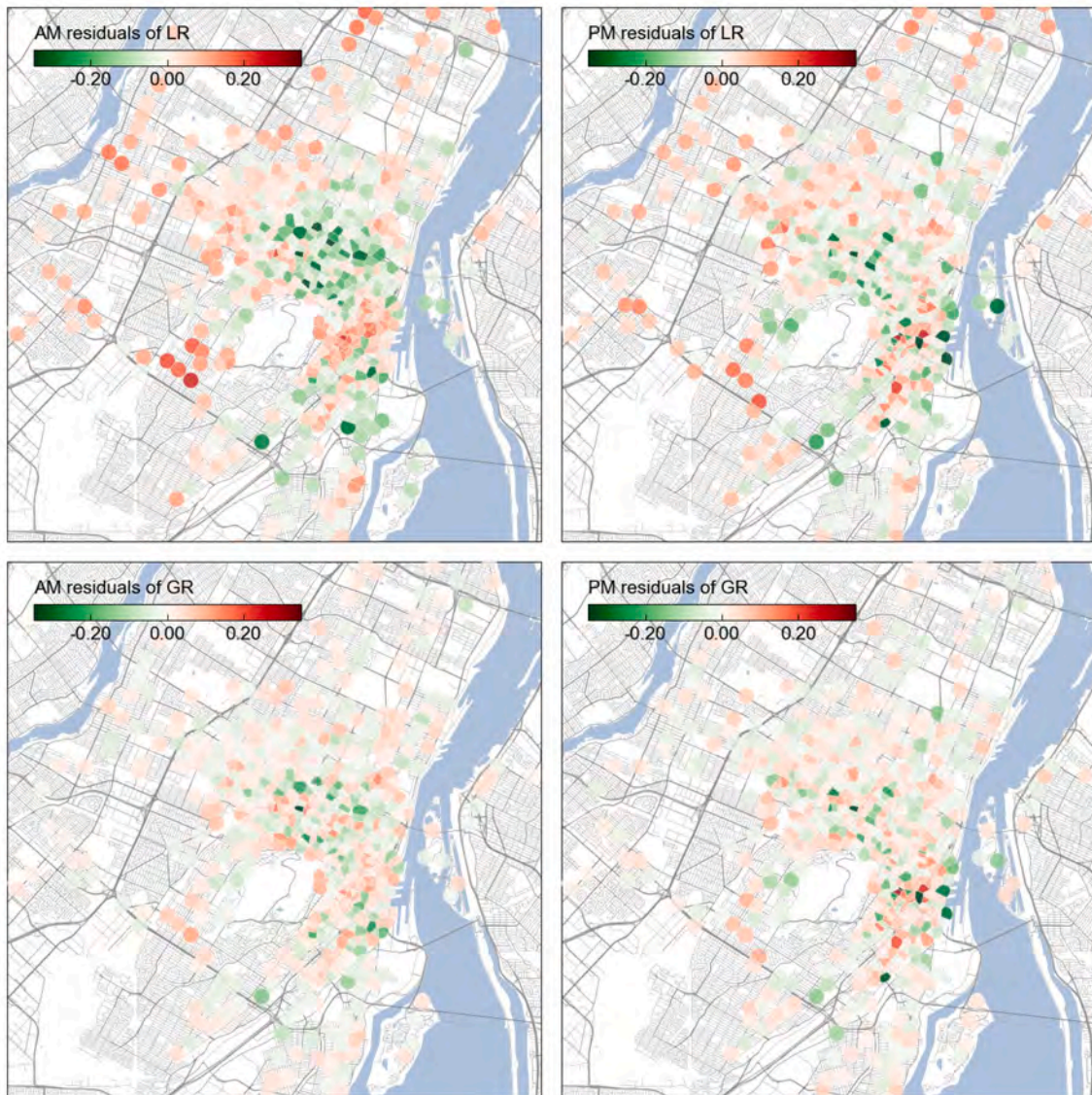
Fig. 9. The regression residuals of LR and GR.

Table 3
Moran's I test on residual.

	Morning		Afternoon	
	LR	GR	LR	GR
Residual sum of squares	6.431	2.570	5.573	3.460
Moran's I	0.340	0.034	0.128	−0.010
p-value of Moran's test	<0.001	0.082	<0.001	0.684

than the real demand for a few "hot spots". While GR is more accurate than LR in general.

We further use Moran's I test (Li et al., 2007) to diagnose the global spatial autocorrelation of the regression residuals. Moran's I is an index ranging from −1 to 1. A random arrangement (no spatial autocorrelation) gives a near-zero Moran's I index. We use the inverse distance weight matrix in the Moran's I test. The test results are shown in Table. 3, where the null hypothesis is that residuals have no spatial autocorrelation. The *p*-values of LR for both morning and afternoon cases are less than 0.001, showing strong spatial autocorrelation. The p-values of GR are larger than 0.05, and we cannot reject the null hypothesis. Therefore, Moran's I test shows the spatial autocorrelation for the

residuals of GR is not significant.

5.4. Out-of-sample prediction

The ability to predict the potential demand of new stations is essential in measuring the goodness of a regression model. It is also of practical significance in predicting the demand for new stations in advance. Examining the out-of-sample prediction also tells if a regression is over-fitted, since GR can infinitely approximate to training data when $\lambda \to 0^+$. Therefore, we evaluate the out-of-sample prediction power of the proposed model by reserving 20% randomly selected stations as a test set. The coefficients of the test set can be estimated based on function (4). To avoid randomness, we repeat the test 20 times with different training and test set separations, and the hyper-parameter setting is fixed for all the tests based on the result of Section 4.2. We use the RMSE in Eq. (5) and the R^2 in Eq. (6) to measure the goodness of the prediction in the test set.

$$R^2 = 1 - \frac{\sum_{i \in \mathcal{V}_{\text{test}}} \left(y_i - \widehat{y}_i\right)^2}{\sum_{i \in \mathcal{V}_{\text{test}}} (y_i - \overline{y})^2}, \tag{6}$$

Table 4

The average RMSE and R^2 with the standard deviation in parentheses in the prediction problem. Best results are highlighted in bold.

	AM departures		PM departures	
	RMSE×10	R^2	RMSE×10	R^2
Linear regression	1.041	0.300	1.016	0.494
	(0.117)	(0.066)	(0.096)	(0.070)
Random forest	0.970	0.392	1.071	0.443
	(0.119)	(0.075)	(0.147)	(0.080)
SVM regression	0.995	0.362	1.053	0.461
	(0.126)	(0.070)	(0.140)	(0.076)
KNN	1.032	0.310	1.124	0.390
	(0.110)	(0.072)	(0.152)	(0.061)
Nearest neighbors average	0.938	0.431	1.209	0.294
	(0.122)	(0.077)	(0.169)	(0.083)
Regression kriging	**0.809**	**0.567**	0.956	0.535
	(0.101)	**(0.050)**	(0.106)	(0.078)
GWR	0.854	0.514	0.959	0.525
	(0.103)	(0.081)	(0.100)	(0.121)
GR (circle buffer)	0.871	0.510	1.014	0.498
	(0.111)	(0.063)	(0.107)	(0.068)
GR (proposed)	0.831	0.554	**0.949**	**0.560**
	(0.104)	(0.051)	**(0.096)**	**(0.059)**

where $\bar{y}$ is the average of the real demand y_i, and a larger R^2 means a higher correlation between the real and the predicted demand.

We compare the proposed model with the following benchmark models in the out-of-sample prediction:

- *Random forest*: randomly construct and merge a multitude of decision trees to predict the demand.
- *SVM regression*: extend support vector machines to solve a regression problem.
- *KNN*: predict the demand based on K nearest factor vectors in the training set.
- *Nearest neighbors average*: the average demand of K nearest neighbors
- *Linear regression*: a global linear regression with invariant factor coefficients.
- *Regression kriging*: apply an ordinary kriging on the residuals of a linear regression model.
- *Geographically weighted regression (GWR)*: a weighted linear regression model with the weight determined by the distance between two observations.
- *Graph regularization with circle buffer*: use 250 m circle buffers to aggregate factors and apply the proposed method.

Random forest, SVM regression, and KNN are classical machine learning models achieved by the scikit-learn package in Python. Geostatistical models, such as Regression kriging and GWR, are completed by R packages automap and GWmodel, respectively. All the hyperparameters of the benchmarks are tuned by a grid search and cross-validation to make a fair comparison. Note we use the same 20 training-test separations for all the models. The code and experiment replication is available in our Github repository.

Table 4 shows the prediction results of the proposed method and benchmarks models. We use the average RMSE and R^2 with the standard deviation in parentheses. Note that we multiply RMSE by ten for a better display effect. Overall, geostatistical models perform better than machine learning models, which indicates the importance of geographic information in predicting the bike-sharing demand. The proposed model is comparable to the two geostatistical models and performs the best in predicting the afternoon demand. The regression kriging exhibits the best performance in predicting bike-sharing demand at morning rush hours. The performance of regression kriging is affected by the spatial correlation of the demand. As shown in Fig. 1, the spatial distribution of the demand in the afternoon is less smooth (weak spatial autocorrelation) than that in the morning. In this situation, the regression kriging is

not as good as the proposed graph regularization method. Also, compared with the proposed model, the factor coefficients in regression kriging are invariant for different stations, which cannot explain the varying effects of a factor to different stations.

The linear regression model shows a relatively poor performance compared with other methods because the linear regression is a global model without considering the spatial heterogeneity of stations. We also apply the proposed graph regularization on the factors extracted from 250-m circle buffers around stations. One shortcoming of the circular buffer is that the catchment areas could overlap, leading to inaccurate factor extraction. As expected, using the catchment areas based on Thiessen polygons has better performance than using circle buffers.

6. Conclusion and discussion

In this paper, we develop a new SVC regression model by graph regularization. This new model can quantify the effect of factors on bike-sharing demand at a station level and predict the demand of new bike-sharing stations. Specifically, we build a linear regression model for each station to obtain its unique factor coefficients and assume the neighboring stations have similar factor coefficients imposed by a graph regularization term. By doing so, the strong spatial correlations/dependencies in bike-sharing can be well-incorporated into the model. Under the same assumption, the potential demand for new stations can also be estimated.

The bike-sharing data collected from BIXI with other urban data are applied in the case study to demonstrate the performance of the model. We divide the BIXI data into the morning peak (AM) and the afternoon peak (PM) to illustrate the temporal characteristic of the factors. From the spatial distributions of factor coefficients, we find that the population and the station capacity are the dominating factors for the bike-sharing demand in the morning and in the afternoon. Specifically, the Pointe-Saint-Charles area shows the largest coefficient of popularity, and the Old Port, a famous sightseeing/commercial area, is highly affected by the bike capacity.

In predicting the demand for new stations, our model outperforms machine learning baseline models and shows comparable performance over geostatistical models (e.g., SAR and SVC). It demonstrates the applicability and potential for the graph-regularization-based SVC models in the spatial regression task.

Besides, the proposed model can also be used for clustering and anomaly detection when applying lasso term $\sum_{(i,j)\in\mathscr{E}} w_{ij}\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|_2$ as the graph regularization (Hallac et al., 2015). The single-element cluster can be regarded as an anomaly since its coefficients are significantly different from its neighbors.

Although the model has various applications, one limitation is that the current model does not consider temporal factors, such as weather, time of day, and cannot work on the short-term prediction task. Another limitation is that the model purely assumes the smoothness of coefficients between nearby stations while it does not fully exploit the dependent variable's spatial correlation. As a result, it is still possible to exist (minor) spatial correlation in the regression residuals.

There are several directions for future research. (1) We define the link weight as the function of the distance between two stations and study the prediction performance with different link weight parameter α. In future work, we can also explore different types of link weight functions, such as considering the correlation between origin and destination (Chai et al., 2018). (2) We can also investigate the effects of temporal factors, such as weather, under a similar framework. For example, turning the factor coefficient $\boldsymbol{\beta}$ into a time-varying version $\boldsymbol{\beta}_t$ by imposing a temporal smoothness assumption. (3) Moreover, the regression model with the graph regularization can be applied to a broader transportation context, such as public transit and urban planning.

Acknowledgment

References

Banerjee, S., Gelfand, A.E., Finley, A.O., Sang, H., 2008. Gaussian predictive process models for large spatial data sets. J. R. Stat. Soc. Ser. B (Statistical Methodology) 70, 825–848.

Bao, J., Shi, X., Zhang, H., 2018. Spatial analysis of bikeshare ridership with smart card and poi data using geographically weighted regression method. IEEE Access 6, 76049–76059.

Brunsdon, C., Fotheringham, S., Charlton, M., 1998. Geographically weighted regression. J. R. Stat. Soc. Ser. D 47, 431–443.

Buck, D., Buehler, R., 2012. Bike lanes and other determinants of capital bikeshare trips. In: 91st Transportation Research Board Annual Meeting.

Cai, S., Long, X., Li, L., Liang, H., Wang, Q., Ding, X., 2019. Determinants of intention and behavior of low carbon commuting through bicycle-sharing in China. J. Clean. Prod. 212, 602–609.

Chai, D., Wang, L., Yang, Q., 2018. Bike flow prediction with multi-graph convolutional networks. In: Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 397–400.

DeMaio, P., 2009. Bike-sharing: history, impacts, models of provision, and future. J. Public Transp. 12, 3.

DMTI Spatial Inc, 2019. Enhanced Point of Interest (DMTI).

Edelsbrunner, H., Seidel, R., 1986. Voronoi diagrams and arrangements. Discrete Comput. Geom. 1, 25–44.

El-Assi, W., Mahmoud, M.S., Habib, K.N., 2017. Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. Transportation 44, 589–613.

Eren, E., Uz, V.E., 2020. A review on bike-sharing: the factors affecting bike-sharing demand. Sustain. Cities Soc. 54, 101882.

Faghih-Imani, A., Eluru, N., 2016. Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: a case study of New York citibike system. J. Transp. Geogr. 54, 218–227.

Faghih-Imani, A., Eluru, N., El-Geneidy, A.M., Rabbat, M., Haq, U., 2014. How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (bixi) in Montreal. J. Transp. Geogr. 41, 306–314.

Fishman, E., 2016. Bikeshare: a review of recent literature. Transp. Rev. 36, 92–113.

Fishman, E., Allan, V., 2019. Bike share. In: Advances in Transport Policy and Planning, 4. Elsevier, pp. 121–152.

Fishman, E., Washington, S., Haworth, N., 2013. Bike share: a synthesis of the literature. Transp. Rev. 33, 148–165.

Gabriel, K.R., Sokal, R.R., 1969. A new statistical approach to geographic variation analysis. Syst. Zool. 18, 259–278.

Gammelli, D., Peled, I., Rodrigues, F., Pacino, D., Kurtaran, H.A., Pereira, F.C., 2020. Estimating latent demand of shared mobility through censored gaussian processes. Transp. Res. C 120. https://doi.org/10.1016/j.trc.2020.102775.

Gelfand, A.E., Kim, H.J., Sirmans, C., Banerjee, S., 2003. Spatial modeling with spatially varying coefficient processes. J. Am. Stat. Assoc. 98, 387–396.

Guidon, S., Reck, D.J., Axhausen, K., 2020. Expanding a (n)(electric) bicycle-sharing system to a new city: prediction of demand with spatial regression and random forests. J. Transp. Geogr. 84, 102692.

Hallac, D., Leskovec, J., Boyd, S., 2015. Network lasso: clustering and optimization in large graphs. In: Proceedings of the 21th ACM SIGKDD International Conference on knowledge discovery and Data Mining, pp. 387–396.

Li, H., Calder, C.A., Cressie, N., 2007. Beyond moran's i: testing for spatial dependence based on the spatial autoregressive model. Geogr. Anal. 39, 357–375.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B (Statistical Methodology) 73, 423–498.

Mateo-Babiano, I., Bean, R., Corcoran, J., Pojani, D., 2016. How does our natural and built environment affect the use of bicycle sharing? Transp. Res. A Policy Pract. 94, 295–307.

Miranda-Moreno, L.F., Nosal, T., 2011. Weather or not to cycle: temporal trends and impact of weather on cycling in an urban environment. Transp. Res. Rec. 2247, 42–52.

Munira, S., Sener, I.N., 2020. A geographically weighted regression model to examine the spatial variation of the socioeconomic and land-use factors associated with strava bike activity in Austin, Texas. J. Transp. Geogr. 88, 102865.

Noland, R.B., Smart, M.J., Guo, Z., 2016. Bikeshare trip generation in New York city. Transp. Res. A Policy Pract. 94, 164–181.

Rixey, R.A., 2013. Station-level forecasting of bikesharing ridership: station network effects in three us systems. Transp. Res. Rec. 2387, 46–55.

Scott, D.M., Ciuro, C., 2019. What factors influence bike share ridership? An investigation of Hamilton, ontario's bike share hubs. Travel Behav. Soc. 16, 50–58.

Shaheen, S.A., Guzman, S., Zhang, H., 2010. Bikesharing in europe, the americas, and asia: past, present, and future. Transp. Res. Rec. 2143, 159–167.

Shen, Y., Zhang, X., Zhao, J., 2018. Understanding the usage of dockless bike sharing in Singapore. Int. J. Sustain. Transp. 12, 686–700.

SIC, 2020. Standard industrial classification (SIC) manual. https://www.osha.gov/data/sic-manual (Accessed: 2021-01-30).

Tobler, W.R., 1970. A computer movie simulating urban growth in the detroit region. Econ. Geogr. 46, 234–240.

Walk Score, 2020. Walk score methodology. https://www.walkscore.com/methodology.shtml. (Accessed 20 July 2020).

Wu, Y., Zhuang, D., Labbe, A., Sun, L., 2020. Inductive graph neural networks for spatiotemporal kriging. arXiv preprint arXiv: 2006.07527.

Yang, F., Ding, F., Qu, X., Ran, B., 2019. Estimating urban shared-bike trips with location-based social networking data. Sustainability 11, 3220.

Yang, H., Zhang, Y., Zhong, L., Zhang, X., Ling, Z., 2020. Exploring spatial variation of bike sharing trip production and attraction: a study based on chicago's divvy system. Appl. Geogr. 115, 102130.

Zhang, Y., Thomas, T., Brussel, M., Van Maarseveen, M., 2017. Exploring the impact of built environment factors on the use of public bikes at bike stations: case study in Zhongshan, China. J. Transp. Geogr. 58, 59–70.