CIRRELT 1971 - 2021

CIRRELT-2020-40

Modeling Bike-Sharing Demand Using a Regression Model with Spatially Varying Coefficients

Xudong Wang
Zhanhong Cheng
Martin Trépanier
Lijun Sun

November 2020

Bureau de Montréal

Université de Montréal
C.P. 6 128, succ. Centre-Ville
Montréal (Québec) H3C 3J7
Tél : 1-514-343-7575
Télécopie : 1-514-343-7121

Bureau de Québec

Université Laval,
2325, rue de la Terrasse
Pavillon Palasis-Prince, local 2415
Québec (Québec) G1V 0A6
Tél : 1-418-656-2073
Télécopie : 1-418-656-2624

Modeling Bike-Sharing Demand Using a Regression Model with Spatially Varying Coefficients

Xudong Wang[1,2], Zhanhong Cheng[1,2], Martin Trépanier[1,3], Lijun Sun[1,2,*]

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT)

[2] Department of Civil Engineering, McGill University, Montreal, QC H3A 0C3, Canada

[3] Department of Mathematics and Industrial Engineering, Polytechnique Montréal

Abstract. Quantifying how different factors affect bike-sharing demand is a critical problem. Most research investigated this problem from a holistic view using regression models, where the coefficients are the same for all bicycle stations. However, a global regression model ignores the local spatial effects of different factors. To address this problem, 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. 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 areas that are more sensitive to the marginal change of a certain factor. The proposed model also exhibits superior out-of-sample prediction power compared with traditional machine learning models and geostatistical models.

Keywords: Bike-sharing system, spatially varying coefficients, spatial prediction, land use and built environment.

Acknowledgements. This research is supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada, Mitacs Canada, and the Canada Foundation for Innovation (CFI). X. Wang would like to thank Fonds de recherche du Quebec - Nature et technologies (FRQNT) for providing scholarship. We also thank BIXI Montreal for providing data. The code for this project and experiment replication is available at https://github.com/mcgill-smart-transport/bikesharing demand land use.

---

* Corresponding author: lijun.sun@mcgill.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2020

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 for 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. Understanding how different factors affect user demand at a station level is important to the planning and operation of bike-sharing service, because bike-sharing systems could be more efficient if planners better identify the incentive/disincentive factors to users (Eren and Uz, 2020), such as weather, built environment and land-use, public transportation, socio-demographic attributes, and temporal factors. More importantly, comprehending the correlations between these factors and user demand can help us estimate the potential demand of new stations in advance, which is a critical step in service planning.

Plenty of research has studied the relationship between station-level bike-sharing demand and various factors by using regression models (Rixey, 2013; Faghih-Imani et al., 2014; El-Assi et al., 2017). However, most existing studies used a single regression model (e.g., multivariate linear regression) with fixed coefficients for all stations, which ignores the spatial heterogeneity of different stations (Bao et al., 2018b). Few studies investigate how these factors impose different effects to different stations, since it is hard to establish a regression model using the data only from one certain station. (Faghih-Imani et al., 2014).

The objective of this study is thus to investigate how influential factors contribute differently to the bike-sharing demand at different stations. In doing so, we establish a regression model with spatially varying coefficients. The dependent variable is the average hourly departure demand at each bicycle station, and we analyze the demand of the morning peak and afternoon peak separately. We select ten factors related to demographics, land-use, transportation infrastructures, and bicycle facilities as independent variables. Each factor is aggregated by catchment areas of stations. Unlike only using the common circle buffers, we use Thiessen polygons (Edelsbrunner and Seidel, 1986) to obtain non-overlapping catchment areas. Inspired by the network lasso problem (Hallac et al., 2015), we design the regression model to be station-specific and regularized by a graph/network structure. The graph regularization in the loss function thus encourages nearby stations to have similar coefficients to each other. The model produces the spatial distribution of the coefficients of different factors, answering which area is more sensitive to the marginal change of a certain factor. In the meanwhile, the factor coefficients and the potential demand of new stations can also be estimated from the known stations. Finally, we demonstrate the superior out-of-sample prediction power of the proposed model by comparing it with traditional machine learning models and geostatistical models.

We summarize the main contributions of this work as follows:

- We analyze the effect of land-use, social-demographic, transportation infrastructure, and bicycle facilities on bike-sharing demand at a station-level.
- We combine circle buffers and Thiessen polygons to extract factors from non-overlapping catchment areas.
- We introduce a graph regularization to a regression problem, working as a spatially varying coefficients model.
- The proposed model can better predict 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 and spatial regression models. 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. Literature Review

2.1. Bike-sharing demand modeling

Over the past decades, there has been a plethora of research analyzing 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 Washinton D.C. by a linear regression model. Rixey (2013) compared the effects of a set 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). Recent research began to consider spatial or temporal dependency of bike-sharing demand. Using spatial lagged features as independent variables, Zhang et al. (2017) studied the spatial correlations in bike-sharing demand and quantified the spatial correlation in demand between nearby stations. Faghih-Imani and Eluru (2016) further incorporated the spatial-temporal interactions into the bike-sharing demand modeling.

In the aforementioned research, only a single regression model was used with invariable coefficients for all stations. However, the effect of a factor can be dissimilar for different stations. To capture the stations' heterogeneity, Hyland et al. (2018) clustered the bike-sharing stations based on types of trips and built separate regression models for each cluster. A limitation of this approach is that it breaks the connection among clusters. The most relevant research to this work is (Bao et al., 2018b), where the authors used geographically weighted regression (GWR) to capture the spatial heterogeneity of the effect of points of interest (POI) to bike-sharing demand. In section 5.2, we will compare the GWR approach with the proposed model.

2.2. Regression with spatially varying coefficients

We introduce a graph regularization to a linear regression model to enable spatially varying coefficients (SVC). There are other SVC models in the literature. The GWR proposed by Brunsdon et al. (1998) is the most commonly used SVC model. It is a weighted local regression model where the weights are determined by the distance between two observations. GWR has been widely used in modeling the demand of taxi (Qian and Ukkusuri, 2015) , public transit (Cardozo et al., 2012), ride-sourcing (Bao et al., 2018a), and bike-sharing (Bao et al., 2018b). Another type of prominent SVC model assumes coefficients follow Gaussian Process (GP). Such as the Bayesian SCV processes (Gelfand et al., 2003), Gaussian predictive processes (Banerjee et al., 2008), and Gaussian Markov random fields using a stochastic partial differential equation link (Lindgren et al., 2011). These models have sound statistical definitions, but the complicated parameter estimation methods limit their applications in large-scale data or models with many SVCs (Dambon et al., 2020). Recent research by Murakami and Griffith (2015) extended eigenvector spatial filtering (also known as Moran's eigenvector mapping) to deal with SVC problem.

The SVC model in this paper is modified from the network lasso problem (Hallac et al., 2015). We establish a graph structure on the bicycle-station network and penalize the coefficients' difference between two nodes connected by an edge. Compared with GP-based SVC models, our approach has a more precise form and can scale to large problems. Moreover, the graph-based approach is more flexible to problems involving non-Euclidean distance (Wu et al., 2020).

3. Data

This section introduces the data applied in the research. Section 3.1 presents the bicycle demand data obtained from BIXI, a bike-sharing system in Montreal. 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 Montreal, BIXI only operates from April to November each year. We use the trip data of 2019 from BIXI to obtain the bicycle ridership, and we only use data between May to October 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 in 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 the factors on bicycle demand 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 meters 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 area on 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 towards 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, the hourly departure in the afternoon 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 could be caused by factors, such as metro stations, bicycle paths, and commercial streets, around the bike stations.

3.2. Influential factors

There are numerous factors, including land-usage, transportation infrastructure, social demographics, and weather condition, 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 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 meters circles and Thiessen polygons (Edelsbrunner and Seidel, 1986) to obtain non-overlapping catchment areas, as shown in Fig. 1. The radius of 250 meters 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 5.2 shows applying Thiessen polygons improves the regression performance compared with circle buffer.

We then obtain independent variables by aggregating factors at each catchment area, as summarized in Table 1. The population is estimated from the demographic data, which is obtained from 2016 Canada census data at a dissemination block level. Land-use factors include the number of commercial points of interest (POI), park, and university. The POI data is obtained from DMTI Enhanced Point of Interest

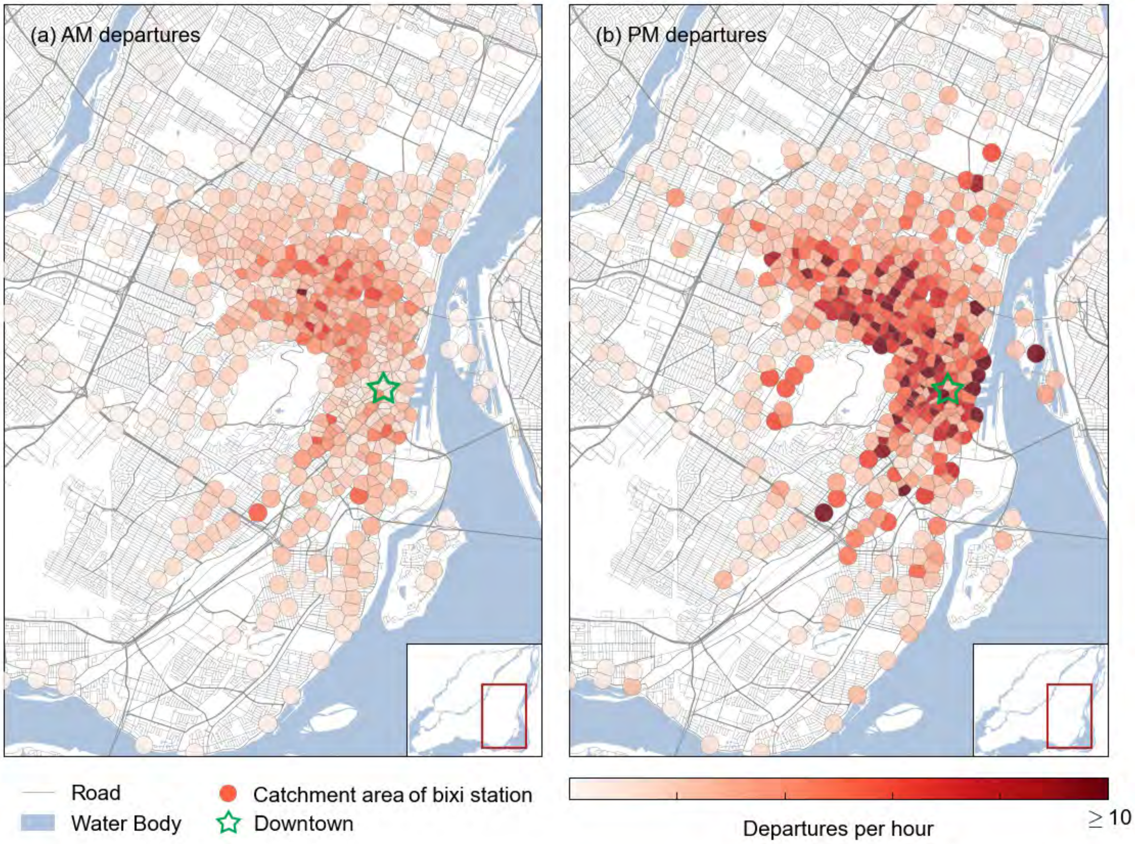

Figure 1: Average hourly departures for BIXI stations in Montreal (left: morning peak, right: afternoon peak).

Table 1: Summary of independent variables (aggregated by catchment areas).

Factors	Description	Min	Max	Mean
Population	Total residential population.	0	4042	1073.88
POI	The number of commercial POI (e.g., wholesale trade).	0	89	14.94
Park	The proportion of parks in the catchment area.	0	1	0.08
University	A binary variable for university.	0	1	0.013
Road length	Total length of roads (in meters).	133.27	4824.40	1934.93
Walkscore	A measure of walkability on a scale from 0 to 100.	7	99	79.28
Metro	A binary variable for metro station.	0	1	0.13
Bus route	log(number of bus routes + 1).	0	3.3	1.281
Cycle path	The proportion of roads with cycle path.	0	1	0.227567
Capacity	The number of docks of a BIXI station.	11	105	23.1

(DMTI Spatial Inc., 2019), we take into account all commercial POI (wholesale trade and retail trade) based on standard industrial classification. Factors related to transportation infrastructures include road length, walkscore (Walk Score, 2020), metro station, and bus route, where walkscore 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. 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.

4. Model

4.1. Regression with graph regularization

Let N be the total number of bicycle stations. For station i ($i \in \{1, \ldots, N\}$), y_i denotes its bike-sharing demand, which is the dependent variable. Denote a vector of independent variables by $\mathbf{x}_i = [1, x_{i1}, \ldots, x_{im}]^\top$ 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:

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

where $\boldsymbol{\beta}_i = [\beta_{i0}, \beta_{i1}, \ldots, \beta_{im}]^\top$ 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:

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

If coefficient vectors $\boldsymbol{\beta}_i$ are the same for all stations, Eq. (1) becomes a simple linear regression model and Eq. (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 Eq. (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. 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 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ are the set of nodes and edges, respectively. Each node represents a bicycle station ($|\mathcal{V}| = N$), 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.

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

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

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, Eq. (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.

Eq. (3) is a convex optimization problem and can be efficiently solved by the commonly used 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

[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.

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:

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

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

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 the optimal values of hyper-parameters. First, all stations are randomly partitioned into 10 equal-sized groups. In each iteration of the cross-validation, retain one group as the validation set $\mathcal{V}_{\text{validate}}$, the remaining stations belong to the training set $\mathcal{V}_{\text{train}}$. For a given combination of hyper-parameters, the coefficients of training stations are estimated by Eq. (3), and the coefficients of test stations are obtained by Eq. (4). Then, the root mean square error (RMSE) is calculated by the demand prediction on the validation set $\mathcal{V}_{\text{validate}}$:

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

where $\hat{y}_i = \mathbf{x}_i^{\top} \boldsymbol{\beta}_i^{\star}$ 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

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 $K^{\star}$

Split the data set $\mathcal{D} = \{X, \mathbf{y}\}$ into 10 equal-sized groups;
for each $\lambda \in \lambda_{set}$, $\alpha \in \alpha_{set}$ and $K \in K_{set}$ **do**
  for each $j = 1, \ldots, 10$ **do**
    $\mathcal{V}_{\text{Validation}}$ = the j th group data of $\mathcal{D}$;
    $\mathcal{V}_{\text{Train}}$ = the remaining data of $\mathcal{D}$;
    Apply CVXPY on $\mathcal{V}_{\text{Train}}$ in Eq. (3) to get $\boldsymbol{\beta}^{\star}_{\text{train}}$;
    Apply CVXPY on $\mathcal{V}_{\text{Validation}}$ in Eq. (4) to get $\boldsymbol{\beta}^{\star}_{\text{validation}}$;
    Get prediction $\hat{\mathbf{y}}$ on $\mathcal{V}_{\text{Validation}}$ in Eq. (1);
    $\epsilon_j \leftarrow$ RMSE in Eq. (5).
  end for
  $E_{\lambda,\alpha,K} \leftarrow \sum_{j=1}^{10} \epsilon_j / 10$.
end for
$\lambda^{\star}$, $\alpha^{\star}, K^{\star} \leftarrow \arg\min\{E_{\lambda,\alpha,K}\}$.
return $\lambda^{\star}$, $\alpha^{\star}$ and $K^{\star}$

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 on the training set is almost 0 because each node is optimized independently. However, apparently the model is over-fitted and generalizes poorly to the validation set. 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 properly 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.

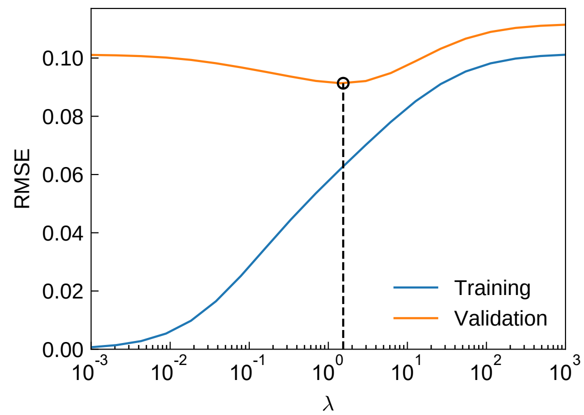

Figure 2: The effect of λ to the RMSE of training and validation set, using demand of afternoon peak, $K = 4$, $\alpha = 1$.

K determines the number of neighbors of a node. When K is very small (1 or 2), the graph may have too many disconnected 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 to 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.

5. Results

5.1. The effect of factors on bike-sharing demand

Using the proposed model, 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. 3, along with a global linear regression model shown by a red line. While the proposed model has varying coefficients, the general result (sign and scale of coefficients) is consistent with the linear regression. Actually, linear regression is a special case of our model when the graph is fully connected and λ is large enough. Note that there are some outliers in the coefficient distributions, which are caused by a small group of disconnected remote stations in the graph, whose bike-sharing demands are very low.

Firstly, we want to clarify the meaning of a coefficient value. It represents how many changes will be in the demand when there is a marginal change in the corresponding factor. For simplicity, we refer to this as

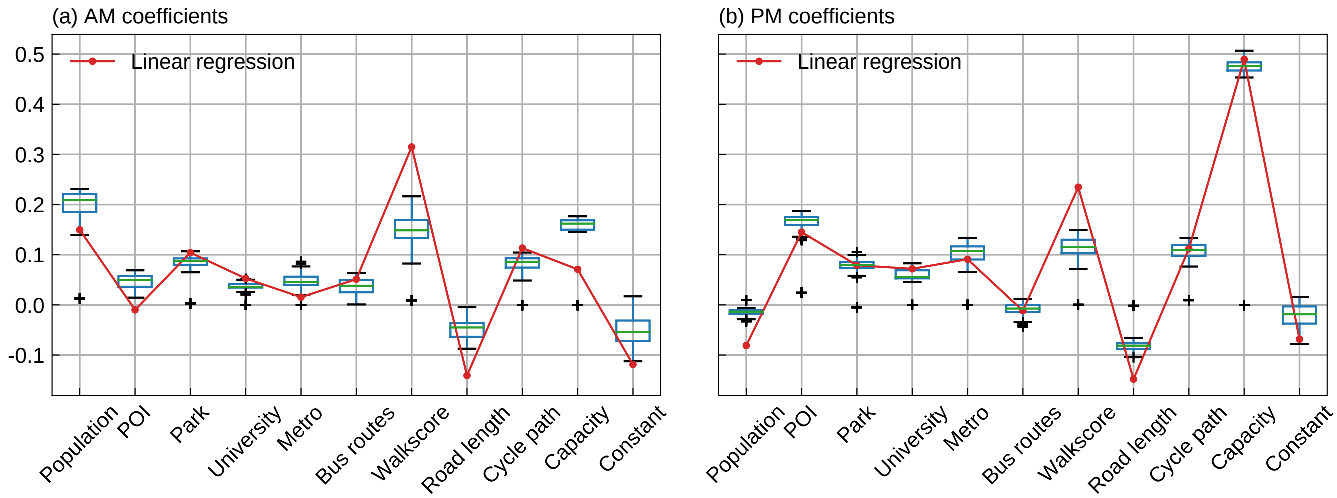

Figure 3: The box plot of factor coefficients for morning peak (left) and afternoon peak (right).

the importance/effect of this factor to the bike-sharing demand. According to Fig. 3, the proposed model shows that the population is the dominating factor for the bike-sharing demand in the morning, while the station capacity has the most dominant effect on the demand in the afternoon. This is easy to understand for the morning since most people depart from their home in the morning. For the afternoon, the relation between demand and capacity is more complex. On the one hand, the demand in the afternoon is much larger than which in the morning (Fig. 1); therefore, a shortage of supply may occur and the real demand could be constrained by the capacity (Gammelli et al., 2020). On the other hand, the station capacity is designed to satisfy the demand pattern, and therefore exhibits a high correlation to the demand in the afternoon peak.

Besides population and capacity, the coefficient of commercial POI 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.

We also visualize the spatial distributions for coefficients in the morning and the afternoon peaks by Fig. 4 and Fig. 5, 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. 3.

In the morning, the Pointe-Saint-Charles area, a lively community with many parks, bike paths, and new housing units, has the largest coefficient of population. In the Plateau-Mont-Royal area and a part of Rosemont-La Petite-Patrie area, POI, bus routes, walkscore are more important. These areas are famous for its commercial streets, delightful parks and attractive culture. Specifically, the 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) has a positive effect on the demand, while the road length has a negative effect. This 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 "unusual stations". For example, there are two unusual points for the metro coefficients shown in Fig. 4, 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. This is because both stations are the transport hub and serve the Orange Line, and the result is consistent with their extremely high metro coefficients and bus route coefficients.

Fig. 5 shows the spatial distributions for factor coefficients in the afternoon peak. Compared with coefficients of the morning, except road length, the number of residential population and bus route also show a negative correlation to the departure demand in the afternoon, which means more these two factors

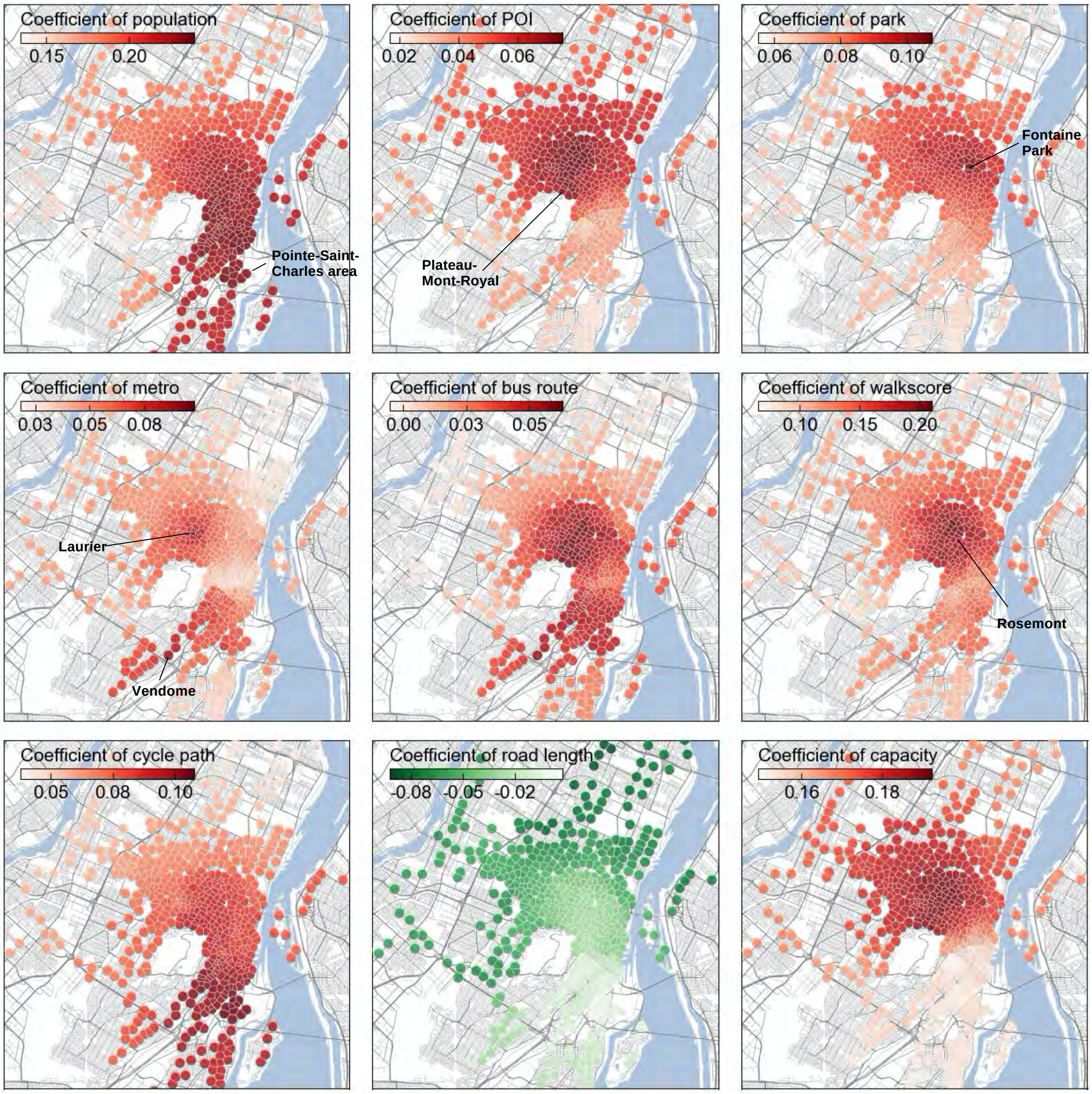

Figure 4: The factor coefficients distribution of morning peak with optimal parameters $\lambda = 2, \alpha = 1$ and $K = 4$.

would have less bicycle demand. This may indicate an inconsistent distribution of the residential and work areas in Montreal. In the afternoon, the POI, walkscore and cycle path still have the great positive impact on 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, etc. The metro station Jean-Drapeau is the main entrance to enter the island except 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 to visit here. So the capacity of bike stations is quite important for the demand at this location.

5.2. 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 of new stations in advance. 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 Eq. (4). To

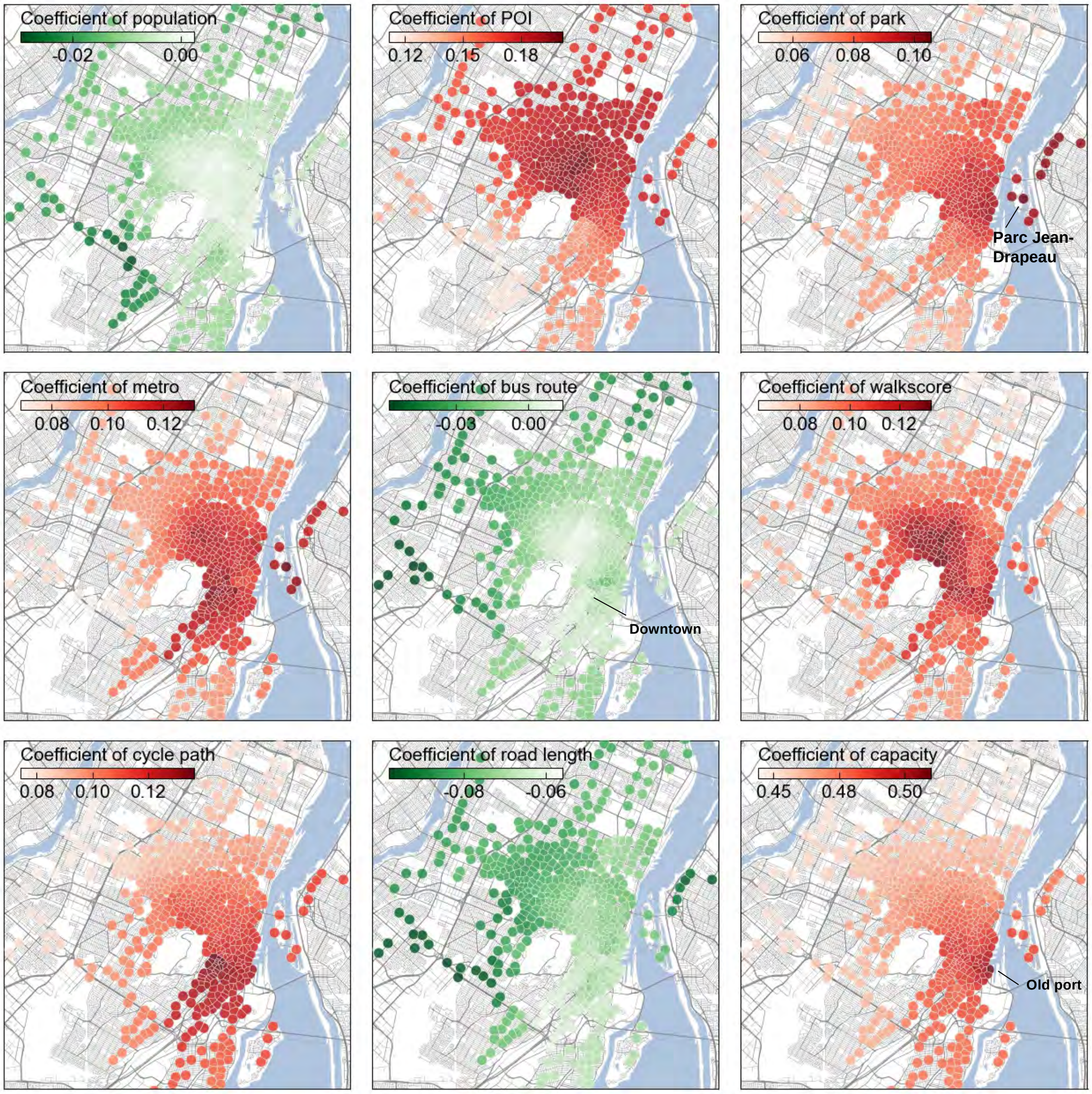

Figure 5: The factor coefficients distribution of afternoon peak with optimal parameters $\lambda = 2.5, \alpha = 1$ and $K = 4$.

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.

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

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 meter circle buffers to aggregate factors and apply the proposed method.

Random forest, SVM regression, and KNN are classical machine learning models; they are achieved by the scikit-learn package in Python and the key hyper-parameters are determined by a grid search and cross-validation. Regression kriging and GWR are geostatistical models; they are achieved by two R packages automap and GWmodel, respectively.

Table 2 shows the prediction results of the proposed method and benchmarks models, where the RMSE and the R^2 are shown by the average value with the standard deviation in parentheses. Note that we multiply RMSE by ten to for a better display effect. Overall, geostatistical models perform better than machine learning models, which indicates the importance of the 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. In addition, 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 meter circle buffers around stations. One shortcoming of the circular buffer is that the catchment areas could overlap and thus leads to inaccurate factor extraction. As expected, using the catchment areas based on Thiessen polygons has better performance than using circle buffers.

Table 2: The RMSE and R^2 in the prediction problem.

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

6. Conclusion and Discussion

In this paper, we develop a regression model with spatially varying coefficients by using a graph regularization. The model can not only quantify the effect of factors on bike-sharing demand at a station level but also 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 of bike-sharing can be well-incorporated into the model. Under the same assumption, the potential demand of unknown stations can also be estimated. In addition, the proposed model can also be used for clustering and anomaly detection when applying lasso term $\sum_{(i,j)\in\mathcal{E}} w_{ij}||\boldsymbol{\beta}_i - \boldsymbol{\beta}_j||_2$ as the graph regularization, in which the single-element cluster can be regarded as an anomaly since its coefficients are significantly different from its neighbors.

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 the population and the station capacity are the dominating factors for the bike-sharing demand in the morning and in the afternoon, respectively. 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. Our method also outperforms other baselines on prediction tasks.

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 thus cannot work on the short-term prediction task. Another limitation is that the model purely assumes the smoothness of coefficients between nearby stations, while does not fully exploit the spatial correlation of the dependent variable itself. As a result, it is still possible to exist (minor) spatial correlation in the regression residuals.

There are several directions for future research. (i) 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 the 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). (ii) We can also investigate the effects of temporal factors, such as weather, under the similar framework. For example, turning the factor coefficient $\boldsymbol{\beta}$ into a time-varying version $\boldsymbol{\beta}_t$ by imposing a temporal smoothness assumption. (iii) Lastly, the regression model with the graph regularization can be applied to a broader transportation context, such as public transit and urban planning.

References

Banerjee, S., Gelfand, A.E., Finley, A.O., Sang, H., 2008. Gaussian predictive process models for large spatial data sets. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70, 825–848.

Bao, J., Liu, P., Blythe, P., Wu, J., 2018a. Exploring contributing factors to the usage of ridesourcing and regular taxi services with high-resolution gps data set, in: Transportation Research Board 97th Annual Meeting, Newcastle University.

Bao, J., Shi, X., Zhang, H., 2018b. 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. Journal of the Royal Statistical Society: Series D (The Statistician) 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. Journal of cleaner production 212, 602–609.

Cardozo, O.D., García-Palomares, J.C., Gutiérrez, J., 2012. Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. Applied Geography 34, 548–558.

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.

Dambon, J.A., Sigrist, F., Furrer, R., 2020. Maximum likelihood estimation of spatially varying coefficient models for large data with an application to real estate price prediction. arXiv preprint arXiv:2001.08089 .

DeMaio, P., 2009. Bike-sharing: History, impacts, models of provision, and future. Journal of public transportation 12, 3.

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

Edelsbrunner, H., Seidel, R., 1986. Voronoi diagrams and arrangements. Discrete & Computational Geometry 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. Sustainable Cities and Society 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. Journal of Transport Geography 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. Journal of Transport Geography 41, 306–314.

Fishman, E., 2016. Bikeshare: A review of recent literature. Transport Reviews 36, 92–113.

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

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

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. arXiv preprint arXiv:2001.07402 .

Gelfand, A.E., Kim, H.J., Sirmans, C., Banerjee, S., 2003. Spatial modeling with spatially varying coefficient processes. Journal of the American Statistical Association 98, 387–396.

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.

Hyland, M., Hong, Z., de Farias Pinto, H.K.R., Chen, Y., 2018. Hybrid cluster-regression approach to model bikeshare station usage. Transportation Research Part A: Policy and Practice 115, 71–89.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73, 423–498.

Murakami, D., Griffith, D.A., 2015. Random effects specifications in eigenvector spatial filtering: a simulation study. Journal of Geographical Systems 17, 311–331.

Qian, X., Ukkusuri, S.V., 2015. Spatial variation of the urban taxi ridership using gps data. Applied Geography 59, 31–42.

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

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

Walk Score, 2020. Walk score methodology. https://www.walkscore.com/methodology.shtml. Accessed: 2020-07-20.

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

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. Journal of transport geography 58, 59–70.