

**Modelling Bicycle Availability in Bicycle Sharing Systems: A Case Study from
Montreal**

Frederic Reynaud

McGill University

817 Sherbrooke Street West, Montreal, Quebec H3A 0C3

Ph: 514-663-3003; Email: frederic.reynaud@mail.mcgill.ca

Ahmadreza Faghih-Imani

Department of Civil Engineering

University of Toronto

35 St George St, GB305, Toronto, ON M5S 1A4

E-mail: a.faghihimani@utoronto.ca

Naveen Eluru* (Corresponding author)

University of Central Florida

12800 Pegasus Drive, Suite 211, Orlando, FL 32816-2450

Ph: 407-823-4815; Email: naveen.eluru@ucf.edu

Modelling Bicycle Availability in Bicycle Sharing Systems: A Case Study from Montreal

ABSTRACT

This paper contributes to the literature on Bicycle Sharing Systems (BSS) by examining bicycle availability at a station as a direct metric of analysis. The main contribution of the current research effort is to develop a behaviorally quantitative model that accommodates for the influence of temporal, meteorological, bicycle infrastructure, built environment and land-use attributes on bicycle availability. An ordered regression model - panel mixed generalized ordered logit model – for hourly bicycle availability is estimated to accommodate for exogenous variables and station level unobserved factors. The model estimation is undertaken using Montreal BIXI data from the summer of 2012. From the results, we observe that BIXI is used more in the afternoon than in the morning, dense areas tend to be associated with lower availability levels, and interactions of time of day with land use impact availability. The estimated model is validated using a hold-out sample. The results clearly highlight the satisfactory performance of the proposed framework. The model developed can be employed by BSS operators to arrive at hourly system state predictions and used for rebalancing operations. To illustrate its applicability, an availability prediction exercise is also undertaken.

Keywords: BIXI, Montreal, bikesharing, Panel Generalized Ordered Logit Model, bicycle availability

1. INTRODUCTION

Bicycle sharing systems (BSS) have been receiving increasing attention in recent years as complementary modes of transportation in urban areas around the world (Parkes et al., 2013). Currently, there are over one million public bicycles worldwide, and over 1,100 cities have installed or are planning a BSS (Meddin & DeMaio, 2015). These systems present many advantages, including flexibility, ease of access and use, physical activity and health-related benefits. These systems also address the issue of bicycle theft for users, a common problem for regular cyclists in urban environments (Van Lierop et al., 2013; Bachand-Marleau et al., 2012). Additionally, BSS offers a potential solution to the “last mile” problem (Cervero et al., 2013; Jäppinen et al., 2013) and are in tune with current generational trends in transportation. Younger generations are less willing to drive, more concerned about the environment, and more prone to use public transit and shared transportation alternatives (Dutzik & Baxandall, 2013).

The observed bicycle flows (arrivals and departures) in a BSS are in response to individuals’ need to travel. Hence, observed flows are significantly influenced by land use and urban form, meteorological and temporal attributes. For example, several studies (Faghih-Imani et al., 2014; Faghih-Imani and Eluru, 2016a; Faghih-Imani and Eluru, 2016b) observed clear commuting trends i.e. in the morning period, bicycles were likely to be picked up from stations farther from the Central Business District (CBD) and dropped off at stations in the CBD. Such asymmetric movements of bicycles in a single direction are likely to create empty stations away from the CBD and full stations around the CBD in the morning hours while the opposite is likely to occur in the PM peak hours. This pattern can lead to lack of access to bicycles (in empty stations) or lack of empty slots (in full stations) for customers. In addition to the commuting trend, several spatial and temporal relationships can result in asymmetry across the system (Nair et al., 2013; Zhao et al., 2015). Such asymmetric usage is a concern for BSS operators because bicycle availability (or empty slot for returning) is at the *heart* of the BSS user-experience. A lack of available bicycles or lack of available space to drop off a bike after usage discourages individuals from using the system. Hence, it is important for system operators to ensure that a desirable level of bicycle availability (or empty slot availability) is maintained. For a fixed station capacity, determining the number of bicycles at the station will automatically determine the number of empty slots. Therefore by examining bicycle availability we automatically observe the availability of empty slots.

To address flow imbalances, in most systems, BSS operators transfer bicycles from full stations to empty stations to ensure bicycle (or slot) accessibility in the system - the process referred to as rebalancing. Moreover, from an environmental perspective, since rebalancing trucks are the only source of air pollution related to BSS systems, it is important to minimize negative environmental externalities. Despite the growth of BSS around the world in recent years and the challenges highlighted above, there are very few studies examining the availability of bicycles or empty slots at a station. To be sure, there have been studies on optimizing rebalancing operations using historical data from a data

mining based approach (Kloimüllner et al., 2014). However, these approaches do not consider any behavioral relationships between BSS demand and factors affecting demand such as socio-demographics, time of day and land use.

In this study, we contribute to the literature on BSS by examining bicycle availability at a station as a direct metric of analysis. With the public availability of trip data on BSS provider websites for various cities (such as Chicago, New York and London), several studies have explored system usage (as arrivals and departures). The approach relates the observed usage (arrivals and departures) at a station to various exogenous factors influencing demand. While system usage is an important measure for BSS operators, it still does not directly study bicycle (or slot) accessibility in the system. The primary reason for lack of analysis of this nature is the lack of station level availability information. The operator provided data does not allow us to generate station level availability because the rebalancing flows - BSS operator undertaken removal or addition - are not reported. The observed availability at the BSS station is a net result of user flows and rebalancing flows. Such information can be obtained through a web crawler based script that records the status of bicycle availability (or slot) in continuous time. In this research effort, we employ a web crawler script based database that directly measures availability to identify potentially full or empty stations.

The current research effort develops a behaviorally quantitative model that accommodates for the influence of temporal, meteorological, bicycle infrastructure, built environment and land-use attributes on bicycle availability. The proposed approach allows system operators to forecast the potential problematic stations (full or empty) based on the quantitative model. Furthering current understanding of the factors affecting bicycle availability will yield insights into the supply-and-demand mechanisms of bikesharing systems, and allow the operators to better optimize their rebalancing procedures and/or plan the system modification (addition or relocation of stations and capacity). In fact, the apriori identification will allow the system operators to plan and schedule rebalancing operations minimizing BSS truck movements while also ensuring improved bicycle (or slot) accessibility in the system. The knowledge of expected problematic stations (full or empty) would also help the BSS operators to formulate innovative programs to reduce the need for rebalancing; for instance, some BSS operators are considering incentive programs for their users to pick up a bike at a full station and drop it off at an empty station (for example see the bike angels program of CitiBike in New York City: <http://bikeangels.citibikenyc.com/>).

The rest of the paper is organized as follows: section 2 gives some background on BSS and reviews the current literature on the topic; section 3 presents the data and models used; section 4 summarizes the results; section 5 features a validation exercise; and finally, section 6 outlines some suggestions for future research and concludes the paper.

2. STUDY IN CONTEXT

2.1. Literature Review

In recent years, several studies have examined bicycle flows and usage levels in various bikesharing systems in Europe and North America. These studies can be segmented into three broad groups. The first group of studies employs actual flow data obtained from the system under consideration to investigate the factors affecting BSS flows. Krykewycz et al. (2010) investigated a planned system in Philadelphia, Pennsylvania, using a raster based Geographic Information System (GIS) to identify possible locations for BSS while using data from European cities to forecast expected demand. Several studies assessed how bikesharing ridership levels were affected by socio-demographic attributes and built environment around stations (Buck and Buehler 2012; Rixey, 2013; Rudloff and Lackner, 2014; Zhao et al., 2014; Wang et al., 2015; Mattson and Godavarthy, 2017). A common limitation of these studies is the lack of detailed temporal resolution. Monthly or annual flow estimations fail to capture short term variation due to shifts in weather, as well as the time of day and weekend variation. Several studies considered fine temporal resolution at an hourly level to examine the influence of meteorological, temporal, land use and built environment attributes on hourly arrival and departure rates (Faghih-Imani et al. 2014; Faghih-Imani and Eluru, 2016a; Faghih-Imani and Eluru, 2016b; Faghih-Imani et al. 2017). Recently, McBain and Caulfield (2018) evaluated estimated travel time differences for ridership trips by comparing observed travel times to the expected travel times. The study found that locations with higher walking access to shops are likely to experience larger differences.

The second set of studies elicits user experience perceptions and system level effects of bicycle-sharing systems. Several studies examined the differences between BSS short-term users and annual members (Lathia et al., 2012; Buck et al., 2013; Faghih-Imani and Eluru, 2015). The convenience of bikeshare systems and the proximity of home to a docking station were found to be the greatest encouragement for individuals to use the system (Fuller et al., 2011; Bachand-Marleau et al., 2012). The relationship between BSS and public transit system was examined in several studies (Bachand-Marleau et al., 2012; Faghih-Imani and Eluru, 2015; González et al., 2015). Fishman et al. (2014) used survey and trip data from five cities to investigate the extent to which BSS can help replace some of the automobile mode share with bicycle share. The study also examined the influence of rebalancing needs in order to determine the impact of BSS on vehicle-kilometers travelled. Bullock et al. (2017) investigated the system costs and benefits associated with BSS. The study found overwhelming evidence of significant economic, time and health benefits from BSS based on their analysis of survey data from Dublin. Godavarthy and Taleqani, (2017) examined willingness to use BSS in the cold season.

Finally, the third group of studies, and the most relevant to current research, focus on identifying problematic stations (stations that are full or empty). Nair et al. (2013) examined system characteristics, utilization patterns, public transit interaction, and flow imbalances between stations over time for the Vélib' system in Paris, France. The authors

adopted a stochastic optimization framework to generate redistribution plans for the Vélib' system. Fricker and Gast (2014), studied the effect of the randomness of user decisions on the number of problematic stations. Vogel and Mattfeld, (2011) and Bouveyron et al. (2015) developed cluster analysis and found different categories of stations within the bikeshare systems. Another subset of studies focuses on operational issues of BSS such as optimizing bicycle repositioning operations and rebalancing truck routing (Vogel and Mattfeld, 2011; Raviv et al., 2013; Kloimüllner et al., 2014; Pfrommer et al., 2014; Forma et al., 2015).

2.2. Study Contribution

The current research effort makes empirical and methodological contributions. Empirically, the study develops a quantitative model to identify stations that are either full or empty, thus providing the system operators with a priori information for optimizing their vehicle routing for ensuring bicycle (slot) availability. As is evident from the literature review, there are few studies exploring the availability aspect in the BSS literature. Earlier work has primarily focused on optimization approaches that work on historical data of BSS usage. While these studies provide useful insights based on analytical approaches, they fail to consider the impact of a host of exogenous variables on bicycle availability. Ignoring the impact of these variables would reduce the effectiveness of the prediction platform for new stations or in locations with rapid land use changes. To elaborate, as these approaches are mainly based on historical patterns, any change in the station structure and usage patterns due to changes in land-use (or new developments) might be harder to replicate.

In our study, we define a new “*availability*” variable measured as the ratio of number of bicycles available to station capacity (ranging from 0 to 1 - empty through full). The variable while continuous in nature does not lend itself to linear regression approaches because of its limited range (between 0 and 1). It is possible to consider a logarithmic form of the variable and employ a log-linear regression approach. While the approach is useful, translating the log-linear variable to meaningful cut-offs for system operators to act upon is not quite intuitive. Hence, in our analysis, we employ an ordered grouping approach that discretizes the availability variable into 5 categories with the following discretization (0-0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8, and 0.8-1.0). The approach provides a directly actionable dependent variable that is easy to consider for BSS operators¹. When the station falls within category 1 (empty) or 5 (full) the station is a potential candidate for rebalancing activity. The proposed availability variable can be employed by any BSS operator in the world to identify stations for rebalancing.

The consideration of the discrete variable presents interesting methodological challenges. The ordered nature of the dependent variable lends itself to an ordered regression approach. However, the traditional ordered outcome models impose a restrictive and monotonic impact – most widely referred to as proportional odds or parallel line

¹ The modelling approach can be readily adjusted for any number of categories.

regression assumption (McCullagh, 1980) – of the exogenous variables on the discrete alternatives. Imposing such restriction can lead to inconsistent parameter estimation. Recent research has offered effective approaches to address the limitation in the form of generalized ordered logit (GOL) models (see Eluru, 2013; Eluru and Yasmin, 2015). These frameworks allow for additional flexibility in modeling the influence of exogenous variables. In our study, the dependent variable of interest – availability - is computed for a station at an hourly level thus offering multiple repetitions of the variable for each station. To account for the influence of unobserved factors affecting availability at a station a panel version of the GOL model needs to be employed. Thus, in our study, a Panel Mixed Generalized Ordered Logit (PMGOL) model is estimated using data from the summer of 2012 to accommodate for exogenous variables and station level unobserved factors. The proposed modeling approach, to the best of the authors’ knowledge, is the first instance of the application of the panel mixed generalized ordered logit model in literature. The estimated model is validated using a hold-out sample of data from the summer of 2013. Finally, to illustrate its applicability, an availability prediction exercise is undertaken. The model developed can be employed by BSS operators to arrive at hourly system state predictions and used for rebalancing operations. In summary, *methodologically*, the paper contributes to research by estimating a panel mixed generalized ordered logit model.

3. DATA PREPARATION AND MODELING EXERCISE

The data used for this study was collected from BIXI Montreal’s website based on the number of bicycles available at each station on a minute-per-minute basis between April and August 2012, for 410 BIXI stations throughout the island of Montreal. The capacity of these stations range from 7 to 65. Around each station, the number of stations within a 250 m buffer vary from 0 to 7. For the purposes of this study, seven consecutive days were randomly selected for each station from the months of May to August in a way that allows for reasonable coverage across the four months. The reader should note that, due to severe winter conditions in Montreal, the BIXI season starts on April 15th. April was excluded from our analysis since all stations are not fully functional. The sample data across each station (as opposed to one week of data for all stations) give us opportunity to better capture the impact of weather attributes such as temperature on BIXI system usage as weather attributes are recognized to be an important factor for cycling (Ahmed et al., 2012; Kashfi et al, 2016). The sampled data was aggregated at an hourly level and augmented with a host of variables, including weather, location, bicycle infrastructure, land use and built environment, and Traffic Analysis Zone (TAZ) level data. The final dataset consists of 68,880 observations (7 days × 24 hours × 410 stations).

3.1. Dependent Variable Definition

An important part of the research exercise is to define station level availability. In our work, we define bicycle availability as the ratio of bicycles docked at a station to station capacity.

Hence, availability of 0 would mean the station is completely empty, while availability of 1 would imply a full station. Further, as BSS operates in continuous time scale, the availability measure could also be computed in continuous time. However, this would make the analysis substantially computationally intensive. Hence, in our approach, we average the minute-by-minute availability across an hour to generate an hourly availability value for each station. Thus, a single hourly measure that reflects the state of the system in that hour is computed as the dependent variable in our analysis. The variable has a range from 0 to 1 with five categories: 0-0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8, and 0.8-1.

Based on the discretization, in 26.3% of cases, stations were less than 20% full, and in 18.3% of cases, they were over 80% full. So, in total the stations were close to unusable 44.6% of the time. These numbers clearly highlight the potential inefficiency in the BSS system being studied. Finally, the spatial distribution of the inefficiency also varies substantially across the system.

3.2. Visual Representation of Availability

In order to better understand bicycle availability and its intimate link with rebalancing operations, we mapped bicycle availability at all stations in Montreal's BIXI system, at 8 AM, 12 PM, 5 PM, and 9 PM of a typical summer day (see Figure 1). The availability values plotted represent the weekly mean of availability values for each time period and station. Figure 1 highlights the typical bicycle movements throughout the day. At 8 AM, stations located near downtown present low availability levels (green), whereas stations located further away are more likely to be full (red). At 12 PM, the trend is reversed. All the morning commutes downtown have filled the downtown stations and emptied the stations located further out. The 5 PM period has less of a clear distinction, since downtown stations are close to empty. Finally, at 9 PM, the downtown stations are nearly empty, while the stations on the outskirts are full. It is interesting to note that there are many more balanced stations (yellow) in the morning than in the evening. This is likely due to the rebalancing efforts of BIXI operators. Mapping the BSS availability across the city clearly highlights the presence of spatial dependencies of bicycle availability at a station with its neighbouring stations. To account for this dependency, we consider a spatial lag parameter as an independent variable in our modeling effort. The spatial lag variable represents the ratio of available bicycles at stations within 250m of a specific station over the capacity of those nearby stations in the previous hour.

3.3. Addressing Rebalancing

In using station data compiled from BIXI's website, it is not possible to differentiate between user drop-offs and pick-ups versus rebalancing actions. Rebalancing operations represent an outside attempt to ensure bicycle availability in the system. However, for our analysis it is critical to account for the presence of artificial flows due to rebalancing. Faghih-Imani et al. (2014) proposed a heuristic approach to separate rebalancing flows

from true flows. However, the same methodology could not be employed here because rebalancing operations are likely to have a more prolonged impact on the dependent variable. To elaborate, accounting for rebalancing on an hourly level would be inadequate, since the number of bikes docked at a station at a specific time is dependent on how many bikes were there in the previous time frame, and will affect several subsequent records. In other words, if a rebalancing operation occurs at 2 pm, it will not only affect the 2-3 pm record. It is much more likely to affect several subsequent records. In order to address this issue, we created identifier variables to recognize rebalancing operations and examined their impact over the next 2, 3, 4, 5, 6, and 12 hours. These identifiers were then provided as input to the models being developed.

3.4. Econometric Model Framework

The Ordered Logit (OL) framework models the choice probabilities of a decision process by partitioning a single latent propensity into as many categories as the dependent variable alternatives through a set of thresholds. Using only a single equation for the propensity, the ordered logit model limits the impact of exogenous variables to be same across all alternatives. To overcome this limitation, Generalized Ordered Logit (GOL) model that allows the exogenous variables to influence the threshold parameters (see Eluru et al. (2008) and Eluru (2013) for more details) was developed. In GOL, in addition to the main equation for the propensity, each threshold can be a function of exogenous variables. Further, to examine bicycle availability at each station for each hour, in this study, we extend the GOL model to a Panel Mixed Generalized Ordered Logit (PMGOL) model that accounts for the panel structure of the data by considering station level unobserved factors. The reader would note that this is the first instance of a panel mixed GOL model implemented in practice.

Consider that propensity for station availability is denoted by y_{it}^* where i represents the station ($i = 1, 2.. N; N=410$ in our case), t represents the hour under consideration ($t = 1, 2.. T; T= 168$ in our case), and j ($j = 1, 2, \dots, J$) denotes the station availability levels. Then, the equation system for PMGOL model can be expressed as follows (Yasmin & Eluru (2013) employed a similar variant for the cross-sectional version of the model):

$$y_{it}^* = (\boldsymbol{\beta} + \boldsymbol{\alpha}_i) \mathbf{X}_{it} + \varepsilon_{it}, \quad (1)$$

and

$$\tau_{it,j} = \tau_{it,j-1} + \exp[(\boldsymbol{\delta}_j + \boldsymbol{\gamma}_{i,j}) \mathbf{Z}_{it,j}] \quad (2)$$

$\boldsymbol{\beta}$ and $\boldsymbol{\delta}_j$ are vectors of unknown parameters to be estimated for exogenous variable matrices \mathbf{X}_{it} and $\mathbf{Z}_{it,j}$ (considered for every threshold) while $\tau_{it,j}$ represents the thresholds associated with these availability levels. In order to ensure the well-defined intervals and

natural ordering of observed availability, the thresholds are assumed to be ascending in order, such that $\tau_{t0} < \tau_{t1} < \dots < \tau_{tj}$ where $\tau_{t0} = -\infty$ and $\tau_{tj} = +\infty$. ε_{it} represents a standard logistic distributed unobserved error term for station i and time period t .

In equations 1 and 2, we assume that α_i and γ_{ij} are independent realizations from normal distribution for this study. Thus, conditional on α_i and γ_{ij} , the probability expressions for station i , hour t and alternative j in MGOL model take the following form:

$$\begin{aligned} \pi_{itj} &= Pr(y_{it} = j | \alpha_i, \gamma_{ij}) \\ &= \Lambda[(\delta_j + \gamma_{it,j}) \mathbf{Z}_{it,j} - (\boldsymbol{\beta} + \alpha_i) \mathbf{X}_{it}] - \Lambda[(\delta_{j-1} + \gamma_{i,j-1}) \mathbf{Z}_{i,j} \\ &\quad - (\boldsymbol{\beta} + \alpha_i) \mathbf{X}_{it}] \end{aligned} \quad (3)$$

where $\Lambda(\cdot)$ represents the standard logistic cumulative distribution function.

The likelihood function conditional on α_i and γ_{ij} , can be written as

$$L | \alpha_i, \gamma_{ij} = \prod_{t=1}^T \prod_{j=1}^J (\pi_{itj})^{d_{itj}} \quad (4)$$

where d_{itj} takes the value of 1 if j is the observed availability at station i for hour t

The unconditional likelihood can subsequently be obtained as:

$$L_n = \int_{\alpha_i, \gamma_{ij}} (L | \alpha_i, \gamma_{ij}) * dF(\alpha_i, \gamma_{ij}) d(\alpha_i, \gamma_{ij}) \quad (5)$$

The log-likelihood function is computed as:

$$\mathcal{L} = \sum_{i=1}^N L_n \quad (6)$$

In this study, we use a quasi-Monte Carlo (QMC) method proposed by Bhat (2001) to draw realization from population multivariate distribution. Within the broad framework of QMC sequences, we specifically use the Halton sequence (250 Halton draws) in the current analysis. The code was programmed in Gauss matrix programming software.

4. ESTIMATION RESULTS

The empirical analysis involves the estimation of three models: (1) the ordered logit (OL) model, (2) the generalized ordered logit (GOL) model, and (3) panel mixed generalized ordered logit model (PMGOL). Prior to discussing the estimation results, we compare the performance of these models in this section based on the Bayesian Information Criterion (BIC) metric. The BIC for a given empirical model is equal to $BIC = -2LL + K \ln(Q)$ where LL is the log-likelihood value at convergence, K is the number of parameters, and Q is the number of observations. The model with the lowest BIC value is the superior model. It is shown in the literature that a reduction of 10 units in BIC is a conclusive improvement in a model for any sample size and number of parameters (Kass and Raftery, 1995). The corresponding values of BIC for the three model frameworks are: (1) OL: 211,403; (2) GOL: 211,039 and (3) PMGOL: 198,001. From the BIC value, it is evident that the

PMGOL with the lowest BIC value substantially outperforms the other variants. This indicates the superiority of considering panel structure and recognizing multiple repetitions of records in model estimation. The better performance of GOL model compared to OL model highlights the positive impact of allowing flexibility in identifying ordinal categories.

The model estimation process was guided by statistical significance (at 90% level), parameter interpretability and parsimony considerations. The panel mixed generalized ordered logit model was estimated by building on the results of the simpler models. The results of the exogenous variable impacts for the PMGOL are presented in Table 1 and are discussed subsequently by variable category. The reader would note that a positive (negative) coefficient value in the propensity equation represents an increase (decrease) in probability of higher level of availability. Similarly, a positive (negative) coefficient value in the threshold equation indicates a shift of the threshold rightward (leftward).

4.1. Constant and Preference Heterogeneity

The constant does not have any substantive interpretation in the model. However, the presence of statistically significant standard deviation on the constant highlights the presence of station specific unobserved effects that influence the availability levels for all records for the station. These common unobserved effects have a standard deviation of 0.4184.

4.2. Weather, Geography and Temporal Variables

The impact of temperature on latent propensity is negative indicating that with increase in temperature, BIXI availability is likely to reduce. This is expected, as in Montreal, with higher temperatures BIXI usage is expected to increase. The coefficient for the elevation variable in the propensity is negative and the coefficient in the third threshold is positive, indicating that stations with a greater elevation are less likely to be full than their counterparts located at lower elevations. As it is easier to bicycle downhill compared to uphill, stations at an elevation are more likely to experience asymmetry in travel to and from such stations.

The results for temporal variables follow expected trends. For instance, the PM coefficient in the propensity function is negative, implying the system is used more in the afternoon than any other time. These results are in line with the findings of Faghieh-Imani et al. (2014). It is noteworthy that the coefficients of AM (6-10am) and PM (3-7pm) are both positive in the second threshold, indicating that stations are more likely to have low availability than to be balanced during those time frames. Overall, since the AM and PM periods are when the system is used most, and the flows are most imbalanced, a concentration of availability around the extremes is expected to occur during those periods. The weekend variable has a positive coefficient, indicating that the system has more availability during the weekdays, due to reduced BIXI usage. To further investigate the

differences in weekdays and weekends, we have tested the interaction of weekend variable with other independent variables to see if their impacts are significantly different during weekends². These estimates are explained where each interacted parameter is discussed in the following sections.

4.3. Bicycle Infrastructure Variables

The number of BIXI stations in a 250-meter buffer offer interesting results. The presence of multiple stations in the 250m buffer is likely to reduce the availability at the station of interest, possibly indicating that these locations are trip generators. On the other hand, in the downtown region, the impact on availability of neighboring stations is compensated by the interaction term, thus indicating that availability is marginally influenced by neighboring stations in the downtown region. The variable interacted with the downtown variable also affects the second threshold, with a negative sign indicating that stations located downtown are more likely to be balanced than low. As expected, a refill rebalancing operation increases availability, while a removal rebalancing operation decreases availability. Further, the estimated coefficient for the influence of availability at neighboring stations in the previous hour is significantly positive as expected; indicating a positive spatial dependency for bicycle availability in a BSS. This result is in agreement with the findings of previous studies on the BSS usage associated spatial correlation (Faghih-Imani and Eluru, 2016a).

4.4. Location, Land Use, and Built Environment Variables

The model indicates that stations located in the old port or downtown areas have lower availability levels overall, which was expected since those are mostly departure areas. During weekends, however, the situation improves slightly in these areas as indicated by positive sign of the interactions with the weekend variable (the overall impact of old port or downtown variables are still negative). Furthermore, these areas are likely to have higher job concentrations and are conducive to PM travel (consistent with findings of Faghih-Imani et al. (2014)). Street length around the station is associated with a positive coefficient in the propensity, which is counterintuitive since one would expect a denser road network in downtown areas. It is important to note that the downtown and old port dummies interact with this variable. Street length is also associated with a negative coefficient in the fourth threshold, indicating that areas with high street length values are more likely to be associated with very high availability.

Walkscore in the vicinity of the station has a negative coefficient indicating highly walkable neighborhoods are bicycle friendly as well. This impact is higher in weekends as highlighted by the coefficient of the interaction with the weekend variable. In addition to

² An interaction variable refers to a new variable created as a product of the variables considered. For example, an interaction variable of weekend and downtown would be computed as weekend variable * downtown variable. The variable allows us to understand how the impact of weekend varies across downtown and non-downtown locations.

the positive mean effect, the Walkscore variable also has a standard deviation indicating that the impact of walkability varies across stations. Further, the propensity function indicates that restaurants affect availability based on time of day. The estimated positive coefficient in the third threshold indicates the lower probability of availability in the AM period in areas with a high number of restaurants. The impact is positive in the PM period showing the presence of restaurants increases availability. This is likely because people usually go to restaurants more in the late afternoon than any other time. Other commercial sites (such as stores, and libraries) exhibit the opposite effect, with a positive impact on propensity in the AM period and negative impact on propensity in the PM period. This suggests BIXI users shop more in the morning than the afternoon. The parameters in the threshold also support the hypotheses for these variables.

4.5. TAZ Level Variables

TAZs with large industrial areas are associated with lower availability levels. This result seems intuitive since industrial parts of town are less likely to be destinations for BIXI users, and unlikely to be refilled. Further, the variable also has a significant standard deviation indicating the impact of the variable varies across stations. In the second threshold, TAZ Parks and Recreational Areas are associated with a positive sign, indicating that stations are more likely to be empty than balanced when they are located in a TAZ with lots of parks and recreational activities. Finally, in the fourth threshold, TAZs with large commercial areas are more likely to have stations with high availability than stations with very high availability. The reasons for this impact are not immediately apparent and warrant further investigation.

5. VALIDATION AND SYSTEM STATE PREDICTION

To evaluate the performance of the PMGOL model, we undertake a validation exercise on a hold-out sample. The same data processing approach is employed for the validation sample preparation. The validation exercise is undertaken at disaggregate and aggregate level. At the disaggregate level, the predictive log-likelihood of the proposed model is estimated. The predictive log-likelihood is compared to the log-likelihood at 0 and log-likelihood at sample shares. The model with 32 parameters show substantial improvements relative to the log-likelihood at 0 and log-likelihood at sample shares. Specifically, the predictive log-likelihood of the PMGOL model is -99,312 while the corresponding numbers for log-likelihood at 0 and at sample shares are -110,858 and -110,088 respectively. The log-likelihood ratio test statistic defined as $(2 \times LL_{UR} - LL_R)$ is computed to evaluate the model fit improvement where LL_{UR} corresponds to the log-likelihood of the unrestricted model (PMGOL model) and LL_R corresponds to the log-likelihood of the restricted model (Model at 0 or Model with constants). The log-likelihood ratio test statistic for our model relative to model at 0 and model with constants are 23,093 and 21,553 respectively. This improvement in predictive log-likelihood is clearly much larger than the

corresponding test statistic for chi-square distribution at any level of significance. Thus, we clearly see that the model predicts the station availability levels adequately.

To undertake comparison at an aggregate level, we compare the predicted aggregate shares with observed aggregate shares. Specifically, we compute the mean absolute percent error (MAPE) value and root mean square error (RMSE) of the predicted shares relative to observed shares. In addition to the full sample comparison, we also examine model performance for two spatial categories: (1) Downtown and Old port and (2) > 5kms from Downtown. The results for the comparison are presented in Table 2. Across all three categories, we observe that the aggregate model performance is very reasonable with MAPE ranging from 2.8% to 7.1%. The RMSE values range from 0.76 to 1.86. Overall, the results indicate high prediction accuracy around the city with slightly lower prediction accuracy near downtown. Even at the downtown with higher usage, the errors are quite satisfactory. Further, we observe a slight over prediction in the very low availability especially in the downtown area based on our model results.

The main strength of the model framework developed is the ability to predict the future availability levels in the bike sharing system. To illustrate this, we provide snapshots of BIXI system availability at 4 instances of the day for a Wednesday in the validation sample. To be sure, the model developed is a probabilistic model and thus only provides the probability of an availability level. To assign the actual availability, we label the alternative with the highest predicted probability as the chosen alternative. A system state prediction based on this approach is presented in Figures 2 and 3. The figures provide evidence of the model's applicability for system state prediction. The figures illustrate the applicability of the proposed model for real-time BSS availability prediction.

BSS operators can plan their rebalancing operations in advance based on expected trends of availability generated by the model. Having a good estimate of BSS's problematic stations and the availability state at each station can help the BSS operators to efficiently preplan their refill and removal operations. If the stations that are most likely to be full or empty in each period are known to the operators, incentive programs can be designed to encourage users to contribute to the system redistribution. In the long term, the model can be re-estimated at regular intervals to allow for the additional feedback associated with the changes in demand and users' behaviour.

6. CONCLUSIONS AND FUTURE WORK

Bicycle sharing systems (BSS) have been receiving increasing attention in recent years as complementary modes of transportation in urban areas around the world. Earlier research exploring BSS has mainly focused on arrivals and departures from a station. The current study addresses this research gap by examining bicycle availability at a station as a direct metric of analysis. Specifically, we estimate an ordered regression model - panel mixed generalized ordered logit model - to accommodate for exogenous variables and station level unobserved factors. Data from Montreal's BIXI system for the summer of 2012 is

employed for model estimation. The model estimation results are intuitive and along expected lines. Specifically, we observe that BIXI is used more in the afternoon than in the morning, dense areas tend to be associated with lower availability levels, and interactions of time of day with land use impact availability. The estimated model is validated using a hold-out sample of data from the summer of 2013. The model validation results clearly highlight the predictive capability of the proposed model. Finally, to illustrate its applicability, we provide system state snapshots for the BIXI system at 4 instances of the day. Such system state prediction serves as useful inputs for undertaking rebalancing exercises.

Future work should investigate the level of data aggregation. The original data was collected on a minute-per-minute basis. This is too fine a resolution for most practical purposes, but whether the data should be aggregated at a 5 minute, 15 minute, half hour, or as we did before at the hourly level, is open to debate and should be investigated further. With the emergence of dockless bikeshare systems with embedded GPS, it might be possible to consider real-time data for rebalancing. Another aspect of interest is the influence of spatial spillover effects from neighboring stations in the system. The models should also consider qualitative indicators such as quality of bicycling and transportation infrastructure in future study efforts. Finally, the predictive models need to be tied to optimization routines to improve routing decision for rebalancing trucks.

ACKNOWLEDGMENTS

The research presented in this report would not have come to fruition without the gracious help of Sabreena Anowar and Shamsunnahar Yasmin. Funding from the National Sciences and Engineering Research Council (NSERC) and the Fonds Québécois de Recherche Nature et Technologies is also acknowledged. The authors would like to acknowledge insightful feedback on an earlier version of the manuscript from three anonymous reviewers and the editor.

REFERENCES

- Ahmed, F., Rose, G., Figliozzi, M., Jakob, C., 2012. Commuter cyclist's sensitivity to changes in weather: insight from two cities with different climatic conditions. Transportation Research Board Annual Meeting, 91st, 2012, Washington, DC, USA.
- Bachand-Marleau, J., Lee, B. H. Y., & El-Geneidy, A. M. (2012). Better Understanding of Factors Influencing Likelihood of Using Shared Bicycle Systems and Frequency of Use. *Transportation Research Record: Journal of the Transportation Research Board*, 2314(-1), 66–71. doi:10.3141/2314-09
- Bhat, C.R. (2001). "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model", *Transportation Research Part B*, Vol. 35, No. 7, pp. 677-693.

- Bouveyron, C., Côme, E., Jacques, J. (2015). The discriminative functional mixture model for a comparative analysis of bike sharing systems. *Annals of Applied Statistics*, Vol. 9, pp. 1726--1760.
- Buck, D., Buehler, R., 2012. Bike lanes and other determinants of capital bikeshare trips. Paper presented at the 91st Transportation Research Board Annual Meeting 2012, Washington, DC.
- Buck, D., Buehler, R., Happ, P., Rawls, B., Chung, P., & Borecki, N. (2013). Are Bikeshare Users Different from Regular Cyclists?: A First Look at Short-Term Users, Annual Members, and Area Cyclists in the Washington, D.C., Region. *Transportation Research Record: Journal of the Transportation Research Board*, 2387(-1), 112–119. doi:10.3141/2387-13
- Bullock, C., Brereton, F. and Bailey, S., 2017. The economic contribution of public bike-share to the sustainability and efficient functioning of cities. *Sustainable cities and society*, 28, pp.76-87
- Cervero, R., Caldwell, B., & Cuellar, J. (2013). Bike-and-ride: build it and they will come. *Journal of Public Transportation*, 16.4, 83-105.
- Dutzik, T., and Baxandall, P. (2013). A new direction: Our changing relationship with driving and the implications for America's future. U.S. PIRG
- Eluru N. (2013). Evaluating Alternate Discrete Choice Frameworks for Modeling Ordinal Discrete Variables, *Accident Analysis & Prevention*, 55 (1), pp. 1-11
- Eluru N., and S. Yasmin (2015) A note on Generalized Ordered Outcome Models, *Analytic Methods in Accident Research* Vol 8, pp. 1-6
- Eluru, N., C.R. Bhat, and D.A. Hensher (2008). A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes, *Accident Analysis & Prevention*, Vol. 40, No.3, pp. 1033-1054
- Faghih-Imani A., N. Eluru, A. El-Geneidy, M. Rabbat, and U. Haq, (2014). How land-use and urban form impact bicycle flows: Evidence from the bicycle-sharing system (BIXI) in Montreal, *Journal of Transport Geography*, Vol. 41, pp. 306-314.
- Faghih-Imani A., and N. Eluru, 2015. Analyzing Bicycle Sharing System User Destination Choice Preferences: An Investigation of Chicago's Divvy System, *Journal of Transport Geography*, Vol. 44, pp. 53-64.
- Faghih-Imani, A., and N. Eluru (2016a), "Incorporating the Impact of Spatio-Temporal Interactions on Bicycle Sharing System Demand: A Case Study of New York CitiBike System," *Journal of Transport Geography*, Vol 54, June 2016, pp. 218-227
- Faghih-Imani, A., and N. Eluru (2016b), "Determining the Role of Bicycle Sharing System Infrastructure Installation Decision on Usage: Case Study of Montreal BIXI System," *Transportation Research Part A* 94C pp. 685-698
- Faghih-Imani, A., Hampshire, R.C., Marla, L. and Eluru, N. (2017). An Empirical Analysis of Bike Sharing Usage and Rebalancing: Evidence from Barcelona and Seville. *Transportation Research Part A: Policy and Practice* Volume 97, Pages 177-191.

- Fishman, E., Washington, S., Haworth, N. (2014). Bike share's impact on car use: Evidence from the United States, Great Britain, and Australia, *Transportation Research Part D: Transport and Environment*, Volume 31, August 2014, Pages 13-20, ISSN 1361-9209
- Fricker, C., & Gast, N. (2014). Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity. *EURO Journal on Transportation and Logistics*, 1–31. doi:10.1007/s13676-014-0053-5
- Fuller, D., L. Gauvin, Y. Kestens, M. Daniel, M. Fournier, P. Morency, and L. Drouin, 2011. Use of a new public bicycle share program in Montreal, Canada. *American journal of preventive medicine*, Vol. 41, No. 1, pp. 80-83.
- Forma, I.A., Raviv, T. and Tzur, M. (2015). A 3-step math heuristic for the static repositioning problem in bike-sharing systems. *Transportation research part B: methodological*, 71, pp.230-247.
- Godavarthy, R.P. and Taleqani, A.R., 2017. Winter bikesharing in US: User willingness, and operator's challenges and best practices. *Sustainable cities and society*, 30, pp.254-262.
- Jäppinen, S., Toivonen, T., & Salonen, M. (2013). Modelling the potential effect of shared bicycles on public transport travel times in Greater Helsinki: An open data approach. *Applied Geography*, Volume 43, Pages 13-24.
- Kashfi, S.A., Bunker, J.M., Yigitcanlar, T., 2016. Modelling and analysing effects of complex seasonality and weather on an area's daily transit ridership rate. *Journal of Transport Geography*, 54, 310–324.
- Kass, R.E. and Raftery, A.E. (1995) Bayes Factors, *Journal of the American Statistical Association*, 90:430, 773-795
- Kloimüller, C., Papazek, P., Hu, B. and Raidl, G.R., (2014). Balancing bicycle sharing systems: an approach for the dynamic case. In *Evolutionary Computation in Combinatorial Optimisation* (pp. 73-84). Springer Berlin Heidelberg.
- Krykewycz, G. R., Puchalsky, C. M., Rocks, J., Bonnette, B., & Jaskiewicz, F. (2010). Defining a Primary Market and Estimating Demand for Major Bicycle-Sharing Program in Philadelphia, Pennsylvania. *Transportation Research Record: Journal of the Transportation Research Board*, 2143(-1), 117–124. doi:10.3141/2143-15
- Lathia, N., S., Ahmed, and L., Capra. Measuring the impact of opening the London shared bicycle scheme to casual users. *Transportation Research Part C: Emerging Technologies*, Vol. 22, 2012, pp. 88-102.
- Mattson, J. and Godavarthy, R., 2017. Bike share in Fargo, North Dakota: Keys to success and factors affecting ridership. *Sustainable cities and society*, 34, pp.174-182.
- McBain, C. and Caulfield, B., 2017. An analysis of the factors influencing journey time variation in the cork public bike system. *Sustainable Cities and Society*.
- Meddin, R. and DeMaio, P., 2015. The Bike-Sharing World Map. Retrieved 13th July, 2015 from <http://www.bikesharingworld.com>

- Nair, R., Miller-Hooks, E., Hampshire, R. C., & Bušić, A. (2013). Large-Scale Vehicle Sharing Systems: Analysis of Vélib'. *International Journal of Sustainable Transportation*, 7(1), 85–106. doi:10.1080/15568318.2012.660115
- Parkes, S. D., Marsden, G., Shaheen, S. A., & Cohen, A. P. (2013). Understanding the diffusion of public bikesharing systems: evidence from Europe and North America. *Journal of Transport Geography*, 31, 94-103.
- Pfrommer, J., Warrington, J., Schildbach, G. and Morari, M., (2014). Dynamic vehicle redistribution and online price incentives in shared mobility systems. *Intelligent Transportation Systems, IEEE Transactions on*, 15(4), pp.1567-1578.
- Raviv, T., Tzur, M. and Forma, I.A., (2013). Static repositioning in a bike-sharing system: models and solution approaches. *EURO Journal on Transportation and Logistics*, 2(3), pp.187-229.
- Rixey, R. A. (2013). Station-Level Forecasting of Bikesharing Ridership: Station Network Effects in Three U.S. Systems. *Transportation Research Record: Journal of the Transportation Research Board*, 2387(-1), 46–55. doi:10.3141/2387-06
- Rudloff, C. and Lackner, B., 2014. Modeling Demand for BikeSharing Systems: Neighboring Stations as Source for Demand and Reason for Structural Breaks. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2430, pp. 1-11.
- Shaheen, S. A., Guzman, S., & Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia: Past, Present, and Future. *Transportation Research Record: Journal of the Transportation Research Board*, 2143(-1), 159–167. doi:10.3141/2143-20
- Van Lierop, D., Grimsrud, M., & El-Geneidy, A. (2013). Breaking into bicycle theft: Insights from Montreal, Canada. *Forthcoming International Journal of Sustainable Transportation*
- Vogel, P. and Mattfeld, D.C., (2011). Strategic and operational planning of bike-sharing systems by data mining—a case study. In *Computational Logistics* (pp. 127-141). Springer Berlin Heidelberg.
- Wang, X., Lindsey, G., Schoner, J., Harrison, A. (2012). Modeling bike share station activity: the effects of nearby businesses and jobs on trips to and from stations. Paper presented at the 92nd Transportation Research Board Annual Meeting 2012, Washington, DC
- Yasmin, S. and N. Eluru (2013), Evaluating Alternate Discrete Outcome Frameworks for Modeling Crash Injury Severity, *Accident Analysis & Prevention*, 59 (1), pp. 506-521
- Zhao, J., Wang, J., & Deng, W. (2015). Exploring bikesharing travel time and trip chain by gender and day of the week. *Transportation Research Part C: Emerging Technologies*, 58, 251-264.
- Zhao, J., W. Deng and Y. Song, 2014. Ridership and effectiveness of bikesharing: The effects of urban features and system characteristics on daily use and turnover rate of public bikes in China. *Transport Policy*, Vol. 35, pp. 253-264.

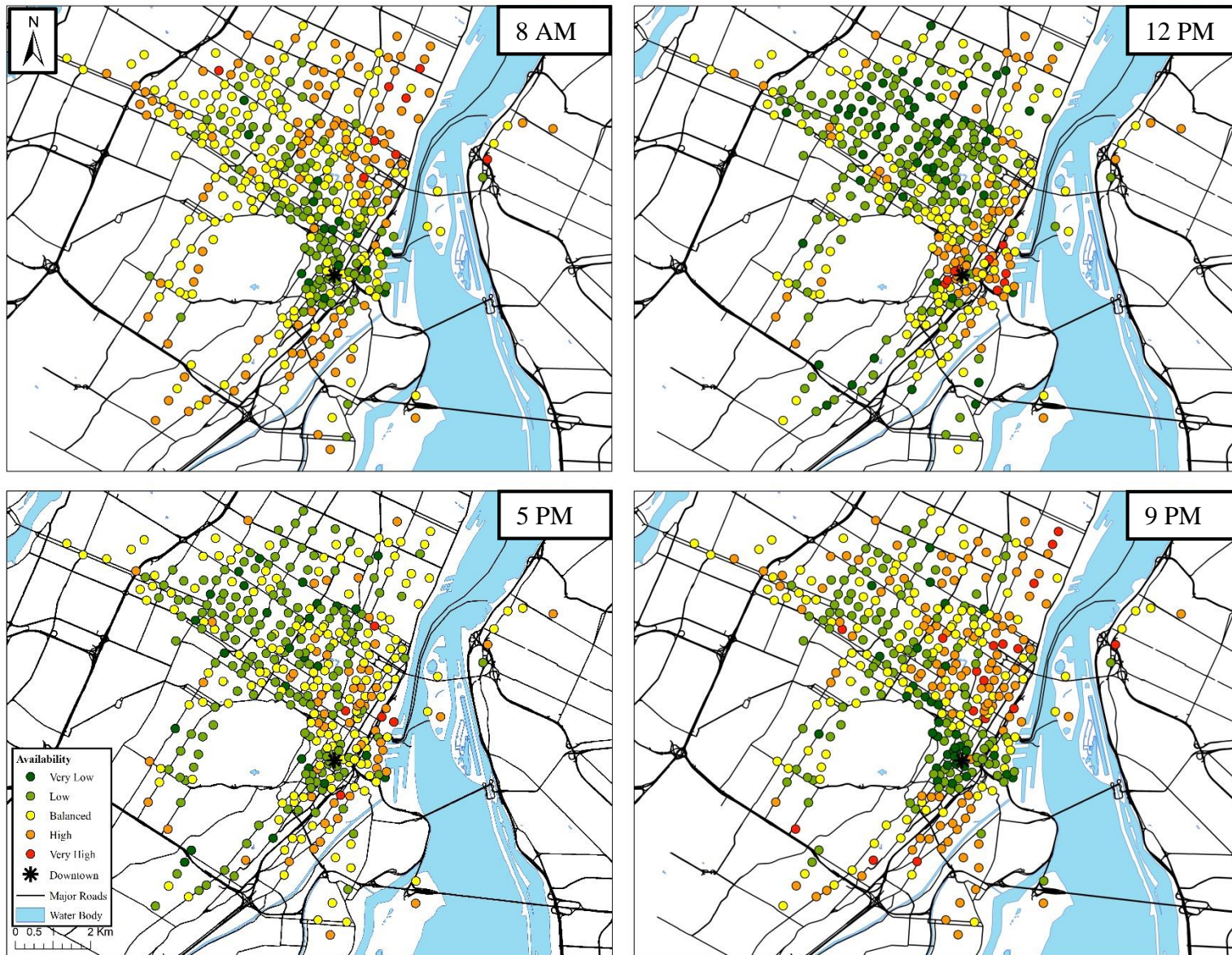


FIGURE 1 Variation of availability during the day around BIXI stations (estimation sample).

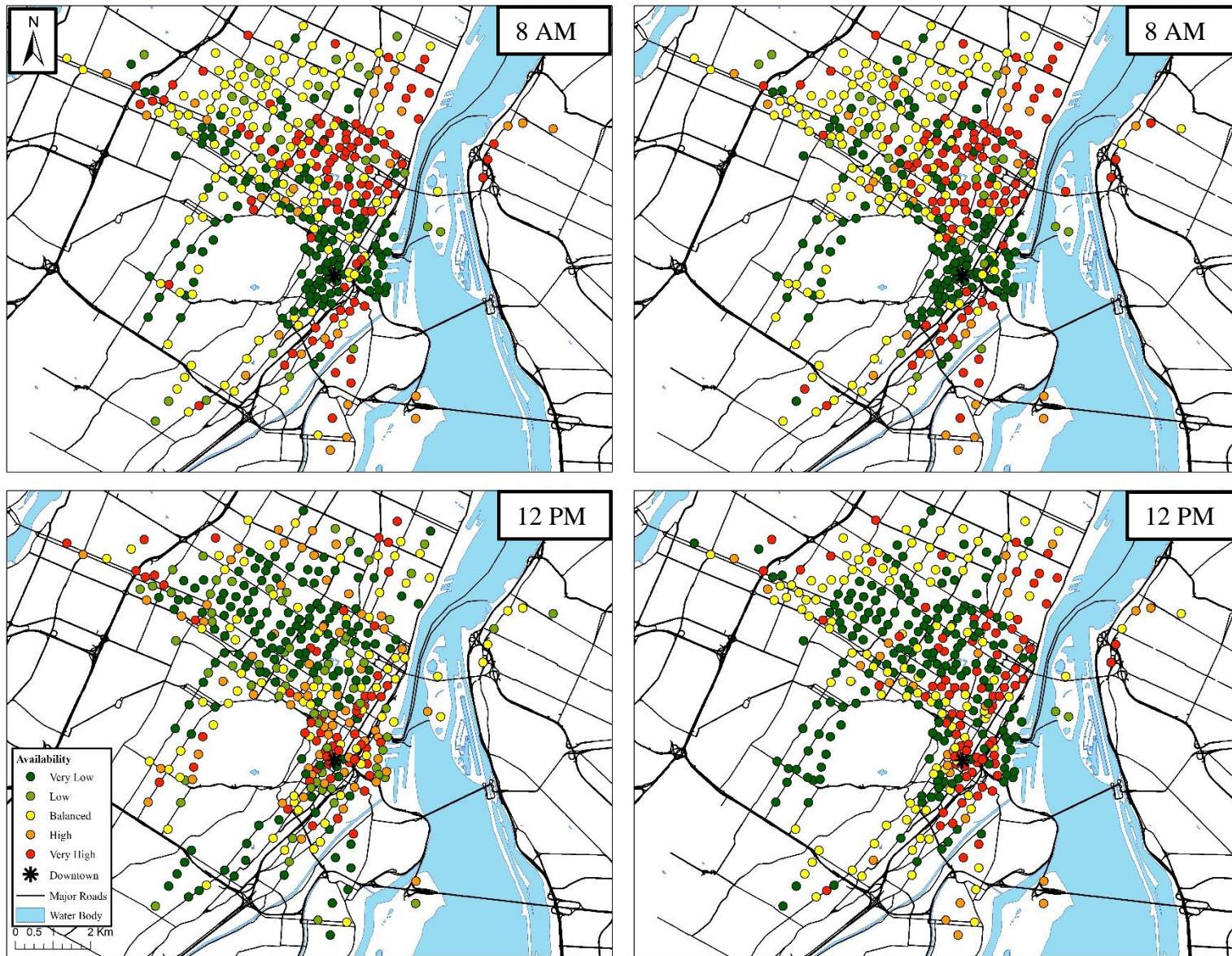


Figure 2 Left: Observed Status, Right: Predicted Status (both on a Wednesday of the validation sample)

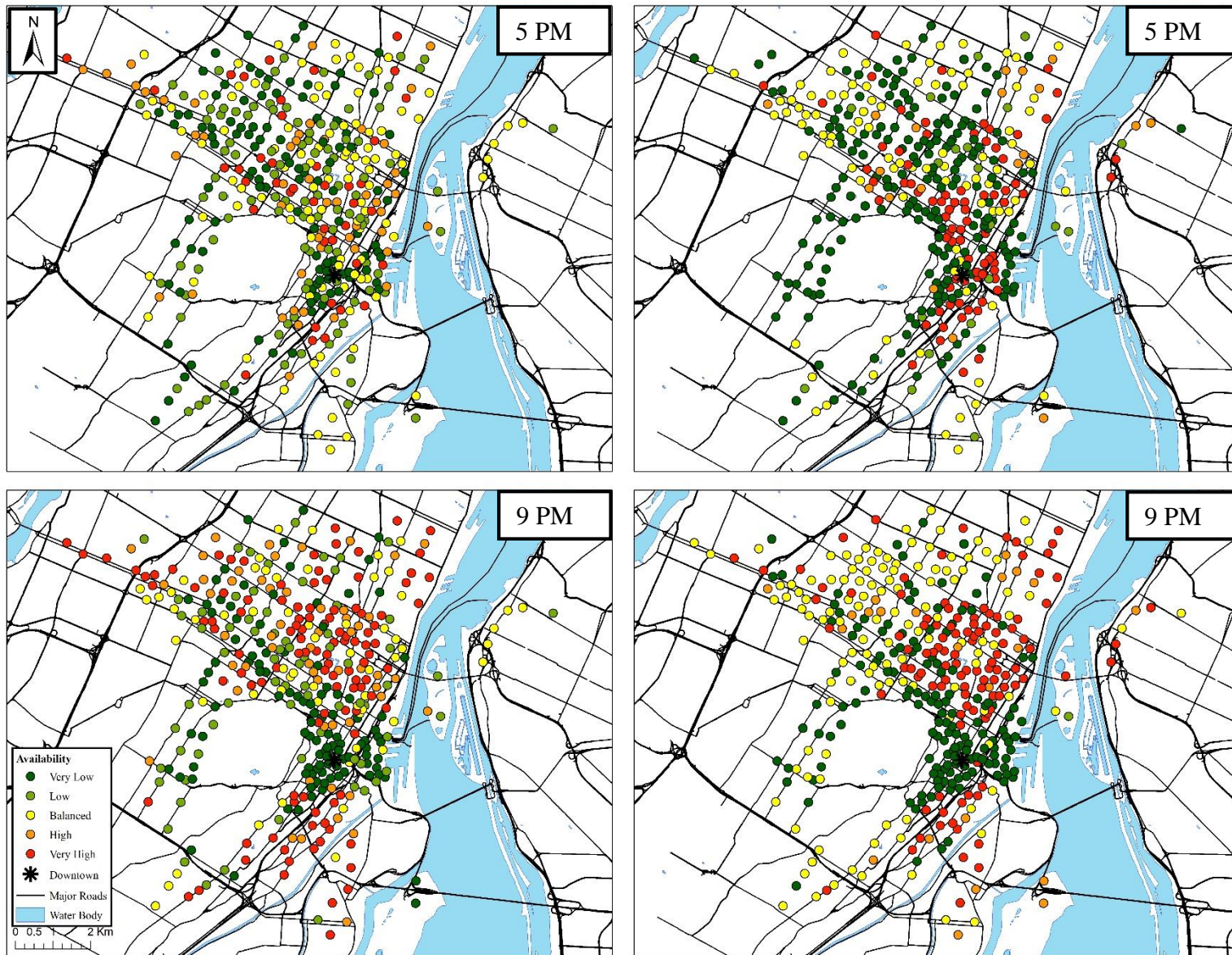


Figure 3 Left: Observed Status, Right: Predicted Status (both on a Wednesday of the validation sample)

TABLE 1 Estimation Results

Variables	Propensity		Threshold b/w Low and Balanced		Threshold b/w Balanced and High		Threshold b/w High and Very High	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Latent propensity component								
Constant	3.170	51.96	0.006	1.11	-0.114	-18.91	0.334	20.68
<i>Standard Deviation</i>	0.418	47.08	-	-	-	-	-	-
Weather, Geography, Temporal								
Temperature (°C)	-0.061	-100.75	-	-	-	-	-	-
Elevation (*10 ⁻¹ ; m)	-0.110	-20.68	-	-	0.026	26.54	-	-
AM period (6-10 am)	-	-	0.098	6.59	-	-	-	-
PM period (3-7 pm)	-0.055	-3.10	0.078	4.97	-	-	-	-
Weekend	0.050	4.15	-	-	-	-	-	-
Bicycle Infrastructure								
Number of BIXI stations in 250m buffer	-1.066	-82.65	-	-	-	-	-	-
Number of BIXI stations in 250m buffer *Downtown	0.674	31.48	-	-	-	-	-	-
Refill (6hr lag)	0.975	20.30	-	-	-	-	-	-
Removal (6hr lag)	-0.355	-6.43	-	-	-	-	-	-
Spatial lag availability (1hr lag)	6.150	313.51	-	-	-	-	-	-
Location, Land use, Built environment								
Old port	-2.081	-44.42	-	-	-	-	-	-
Old port interacted with Weekend	0.211	9.66	-	-	-	-	-	-
Downtown	-2.644	-29.86	-	-	-	-	-	-
Downtown interacted with Weekend	0.148	10.25	-	-	-	-	-	-
Street length in 250m buffer (km)	0.479	36.11	-	-	-	-	-0.094	-20.74
Walkscore (1: low - 7: high ; *10 ⁻¹)	-0.015	-2.81	-	-	-	-	-	-
<i>Standard Deviation</i>	0.245	79.55	-	-	-	-	-	-
Walkscore (1: low - 7: high ; *10 ⁻¹) interacted with Weekend	-0.005	-2.39	-	-	-	-	-	-
Restaurants in 250m buffer interacted with AM (*10 ⁻²)	-	-	-	-	0.463	7.70	-	-
Restaurants in 250m buffer interacted with PM (*10 ⁻²)	0.257	5.11	-	-	-	-	-	-
Commercial venues in 250m interacted with AM (*10 ⁻³)	0.558	6.68	-	-	-0.642	-5.21	-	-
Commercial venues in 250m interacted with PM (*10 ⁻³)	-0.515	-5.80	-	-	-	-	-	-
TAZ Industrial and Resources (km ²)	-4.361	-21.68	-	-	-	-	-	-
TAZ Parks and Recreational Areas (km ²)	-	-	0.124	4.76	-	-	-	-
TAZ Commerces (km ²)	-	-	-	-	-	-	8.895	53.82
Log-likelihood at convergence (Number of observations)	-98,811 (68,880)		- = Not applicable					

TABLE 2 Aggregate Measures of Fit

Availability levels/ Measures of fit	Full sample		Old port and Downtown		>5 km from Downtown (not old port)	
	Actual shares (% records)	Predicted shares (% records)	Actual shares (% records)	Predicted shares (% records)	Actual shares (% records)	Predicted shares (% records)
0-0.2	25.8	26.8	36.8	39.4	19.5	19.8
0.2-0.4	18.1	18.4	17.3	18.2	17.5	19.4
0.4-0.6	20.4	19.5	16.2	16.1	24.9	22.7
0.6-0.8	17.4	16.8	13.4	12.1	20.6	20.2
0.8-1	18.3	18.5	16.4	14.2	17.6	18.0
MAPE	2.84		7.12		5.08	
RMSE	0.76		1.86		1.49	
Records	68,880		15,120		16,632	