Real-time Prediction of Public Bike Sharing System Demand using Generalized Extreme Value Count Model

Soheil Sohrabi
Civil & Environmental Engineering Department
Old Dominion University
Norfolk, VA 23529
Tel: 757 683 5670
Email: ssohr001@odu.edu

Rajesh Paleti (Corresponding Author)
Civil & Environmental Engineering Department
The Pennsylvania State University
University Park, PA 16802
Tel: 814 863 4291
Email: rpaleti@psu.edu

Lacramioara Balan
Civil & Environmental Engineering Department
Old Dominion University
Norfolk, VA 23529
Tel: 757 683 5670
Email: lbala001@odu.edu

Mecit Cetin
Civil & Environmental Engineering Department
Old Dominion University
Norfolk, VA 23529
Tel: 757 683 6700
Email: mcetin@odu.edu

Date of Submission: 10/09/2017
Date of First Revision: 10/02/2019

Abstract

Public Bike Sharing Systems (BSSs) are becoming increasingly popular in recent times. Both the BSS operators and the customers can benefit from the large digital data portals that continuously record the state of the BSS. In this context, the current study developed generalized extreme value (GEV) count models that can predict hourly bike arrivals and departures at each station while accounting for time-of-day, weather, built environment, infrastructure, temporal, and spatial dependency factors. The proposed models were used to analyze the demand patterns in the Capital Bikeshare system and were found to predict the demand at both aggregate and disaggregate levels with reasonable accuracy. Specifically, the total demand in the entire system was predicted within 5% margin of error whereas 75% of the station-level arrival and departure predictions in the next one hour were within a margin of one from the observed counts. The proposed modeling system is useful (a) to BSS customers to better plan their travel based on expected bike and dock availability at the origin and destination ends of their BSS trips, and (b) to BSS operators to anticipate the future demand and optimize their rebalancing plans.

***Keywords*:** Bike sharing, real-time prediction, General extreme value models; count data

1. Introduction

Active or non-motorized transportation that includes walking and cycling has received considerable attention recently due to the associated health, environmental, and transport benefits. For instance, mitigating urban traffic congestion levels, reducing greenhouse gas (GHG) emissions, lower transportation costs, and flexible mobility and support for multimodal transport connections are some of these benefits offered by active transportation modes (Gotschi and Mills, 2008). Understandably, many cities across the world have begun investing in building new and expanding existing non-motorized infrastructure systems (Shaheen et al., 2010, Piatkowski et al., 2015). Public Bike Sharing Systems (BSSs) are an excellent example of such recent efforts to promote active transportation. The current version of BSS is shaped after a long trial-and-error process that evaluated free bikes, coin-deposit systems, IT-based systems with docking stations and check-in check-out policy, and demand-responsive multimodal systems that integrate several transportation modes (Shaheen et al., 2010). To date, this evolution has not stopped and emerging trends indicate effective use of emerging technologies including dockless systems and transit smartcard integration (Parkes et al., 2013). It is essential to optimize the performance of the BSSs to ensure that they remain competitive and viable modal alternatives. The performance of a BSS system can be defined in terms of (a) travelers' satisfaction including accessibility, convenience, safety, and security and (b) the monetary profits earned from the health, environmental, social and transportation benefits. Adopting new technologies to facilitate user access to bikes and expediting the check in/out process, constructing appropriate trails, expanding the system network, integrating other transportation modes with the system, increasing the reliability of the system by providing sufficient amount of bikes are some ways to meet the user's satisfaction (Shaheen et al., 2010, Parkes et al., 2013, Shaheen et al., 2013).

The first step towards making BSS investment decisions is prediction of the BSS use (i.e., demand) under different policy scenarios as a function of travelers' socio-demographics, land-use and built-environment attributes, temporal, and weather conditions (Froehlich et al., 2009, Handy et al., 2010, Faghih-Imani et al., 2014). These studies can be categorized into two groups based on the time-horizon of prediction–long-term and short-term. Long-term predictions unveil the needs of pre-launch planning for developing the BSS system. Also, for BSS systems already in operation, sensitivity analysis of weather, time-of-day, day-of-week, seasonal, land-use, accessibility of facilities or other modes of transportation, and special events on BSS demand can provide the BSS operator insights into the maintenance/expansion scheduling. Short-term prediction helps system operators to improve day-to-day operations to enhance the reliability and accessibility of the system. The temporal and spatial demand patterns have significant implications for BSS operation, the most important one being that the system will get skewed if the BSS operator does not intervene and redistribute bikes periodically. For instance, in the morning, all commuters will pick up bikes at stations in residential zones and travel to work zones leaving very few bikes for other non-working customers in the off-peak hours. So, the BSS operator must intervene and rebalance the bikes and empty docks to meet the expected demand patterns. Currently, rebalancing schedules are either based on historical trends or simple heuristics (e.g., all stations must be half empty). However, optimal rebalancing strategies that minimize the disutility associated with not finding a bike or an empty slot are missing.

This problem can be addressed by drawing strength from BSS digital data portals that provide large streams of real-time system state data. In this context, the objective of the current study is to develop real-time bike share demand forecasting models that predict the number of departures of customers retrieving bikes and number of arrivals of customers returning bikes at each station by time-of-day. However, BSS data has several unique characteristics that cannot be handled using standard count models available in the literature. On the other hand, incorporating additional features to improve model flexibility comes at the cost of increased model complexity and reduced practical applicability. The current study addresses these challenges by developing a computationally tractable count modeling framework that accounts for (1) over or under-representation of multiple arrival or departure count outcomes, (2) station-specific random effects, (3) temporal dynamics, and (4) spatial dependencies. The study accomplishes these research objectives by re-casting standard count models as special cases of Generalized Extreme Value (GEV) models that have seen wide applicability in transportation given their closed-form probability expression. The resulting model system not only enhances accessibility to customers by enabling them to plan their trips efficiently (*e.g.*, lower wait times for returning/retrieving bikes) (Kaltenbrunner et al., 2010), but also assists BSS operators by enabling dynamic rebalancing schedules based on expected demand patterns (Contardo et al., 2012, Ghosh et al., 2017). The applicability of the proposed modeling system was demonstrated for the Capital Bikeshare system that is one of the oldest and largest BSS operating in Washington D.C., Virginia, and Maryland.

The rest of the paper is structured as follows. First, a brief overview of past literature on BSS demand prediction is presented. Next, a comprehensive description of the methodological framework is provided. This is followed by sections describing data, empirical results, and evaluation of prediction accuracy of the proposed models. Then, we elaborate on the practical applicability of the proposed model. Lastly, the paper concludes with potential future extensions.

2. Literature review

Previous studies on cycling demand prediction share a common objective in that they correlate the demand to temporal, meteorological, geographic, and socio-demographic factors. However, these studies differ in (1) the time-horizon and spatial resolution, (2) the modeling methodology and (3) the explanatory variables considered.

The time-horizon for predictive modeling varies between monthly (Rixey, 2013), weekly (Schneider et al., 2009), daily (Lindsey et al., 2007, Lindsey et al., 2008, Wang et al., 2016), and hourly (Kaltenbrunner et al., 2010, Jones et al., 2010, Gallop et al., 2011, Faghih-Imani et al., 2014, Faghih-Imani et al., 2017) time scales depending on research context. However, the literature suggests that bigger time windows will lead to less accurate real-time forecasts (Yoon et al., 2012). With regard to spatial resolution, cycling demand was modeled for the city as a whole (Gallop et al., 2011, Ermagun et al., 2017), for clusters of stations (Li et al., 2015, Liu et al., 2016), and for each bike station (Kaltenbrunner et al., 2010, Rixey, 2013, Faghih-Imani et al., 2014).

The prediction method adopted varied based on the temporal and spatial resolution of analysis and the demand metric used in the study. For aggregate demand modeling over larger geographic scales and longer time-horizons, the linear regression model and its variants were used (Lindsey et al., 2007, Lindsey et al., 2008, Flynn et al., 2012). Count models, specifically Poisson and Negative Binomial (NB) models, with the ability to generate non-negative integer predictions, were used for station-specific and short-term forecasting (e.g., daily and hourly demand modeling) (Wang et al., 2013, Wang et al., 2016). Mixing models that account for station-specific random effects (Faghih-Imani et al., 2014, Faghih-Imani et al., 2017) and time-series models that account for temporal dynamics (i.e., the effect of demand in earlier time-periods) were also developed. For instance, time-series auto-regressive moving average (ARMA) modeling method was used for predicting bike availability at stations few minutes into the future (Kaltenbrunner et al., 2010). Pattern recognition and machine learning techniques are adopted as opposed to econometric models for short-term demand prediction (Li et al., 2015, Liu et al., 2016).

A wide variety of explanatory variables were considered in past research and these can be grouped under five categories–socio-demographics, built environment, infrastructure, weather, and temporal or time-related. Socio-demographic variables capture the effect of potential users on BSS demand. Age, gender, marital status, employment status, income, the number of students and race are examples of socio-demographic factors used in the past (Moudon et al., 2005, Wang et al., 2016). Land-use or built environment attributes describe the neighborhood at the origin and destination ends of bike trips. Examples include park area, shopping, and recreational opportunities, and employment density. The infrastructure variables describe the BSS infrastructure (e.g., the number of stations and their capacity), other bike infrastructure (e.g., length of bike trail in the vicinity of stations), and other transportation infrastructure (e.g., the number of metro and bus stations) (Schneider et al., 2009, Jones et al., 2010, Schneider et al., 2012, Faghih-Imani et al., 2014, El-Assi et al., 2017). In a survey of 1,402 current and potential cyclists in Metro Vancouver, 73 motivators and deterrents for cycling were evaluated (Winters et al., 2011). The findings show that weather was one of the key determinants of cycling. Earlier studies used precipitation, humidity, temperature, and wind speed for predictive modeling. The non-linear effect of weather variables was captured by using indicator variable specifications (Phung and Rose, 2007, Ahmed et al., 2010, Lewin, 2011, Miranda-Moreno and Nosal, 2011) or a psychological temperature perception variable (Böcker and Thorsson, 2014). Temporal factors including day-of-week, holidays, seasons, time-of-day are ubiquitous in BSS demand models.

3. Methodology

The number of bike arrivals and departures at a bike station in an hour is a count response variables. Parametric count models including Poisson and Negative Binomial (NB) models (and their variants) have served as the standard workhorse models for analyzing count data. The Poisson model is suited for count data with the *equi-dispersion* property, *i.e.* the mean is equal to the variance. However, in many empirical scenarios, count data tends to be over-dispersed, *i.e.* the mean is less than the variance. In such cases, the NB model that may be viewed as a Poisson model with gamma distributed mean parameter is better suited (Greene, 2008). It is also common to see that the zero outcomes are over-represented in count data, *i.e.* the proportion of observations with zero outcomes is more than those predicted by typical parametric count models. Researchers used either the hurdle or zero-inflated modeling framework to handle this '*excess zeroes*' problem (Gurmu, 1998, Lord et al., 2005). However, extending these hurdle and inflated frameworks to handle over or under representation of multiple count outcomes (*i.e.,* excess ones, excess twos, *etc.*) leads to a complicated model structure that is not parsimonious and difficult to estimate. Recently, Generalized Extreme Value (GEV) count models that can easily handle probability mass deviations of multiple count outcomes from parametric count models were developed. Moreover, these GEV count models subsume the standard count models (including Poisson and NB) as special cases (Paleti, 2016). Each station has data corresponding to multiple days and multiple time-periods. It is likely that there are station-specific unobserved factors that influence bike arrivals/departures across all days and time-periods. These common unobserved factors that remain same across all stations for all days can be captured by introducing station-specific random effects into the mean parameter of the count model. In the current empirical context, temporal and spatial dynamics are key to determining the future system state. Specifically, the number of bike arrivals/departures in the previous time-periods and neighboring stations influence the number of bike arrivals/departures in the next time-period of proximal stations (Narayanamoorthy et al., 2013).

Let $s(=1,2,\ldots S)$ and $t(=1,2,\ldots T)$ denote the index for bike station and time-period, respectively. In the GEV Poisson model framework, the propensity associated with count outcome i can be written as:

$$U_{s,t,i} = \tilde{V}_{s,t,i} + V_{s,t,i} + \varepsilon_{s,t,i} = ln\left(\frac{\lambda_{s,t}^{i}}{i!}\right) + V_{s,t,i} + \varepsilon_{s,t,i} \qquad \text{Equation (1)}$$

The $\varepsilon_{s,t,i}$ term is the stochastic component of the propensity function. The $V_{s,t,i}$ term is the additional propensity component specific to count outcome i to provide more flexibility over and beyond simple Poisson count model. All count alternatives beyond a certain value J may be considered as the base alternative group during the specification of these additional propensity terms, *i.e.*, $V_{s,t,j} = 0 \; \forall \; j \geq J$. This $V_{s,t,i}$ term takes the form of linear-in-parameters specification as follows:

$$V_{s,t,i} = \boldsymbol{\delta}_i' \boldsymbol{z}_{s,t,i} \qquad \text{Equation (2)}$$

where $\boldsymbol{z}_{s,t,i}$ is the vector observed attributes and $\boldsymbol{\delta}_i$ is the corresponding vector of parameters.

The $\lambda_{s,t}$ is the expected count frequency at station s in time-period t and is parameterized as:

$$ln(\lambda_{s,t}) = \sum_{k=1}^{K} \beta_k x_{s,t}^k + \sum_{r=1}^{t-1} \alpha_r y_{s,r} + \sum_{k=1}^{K} \gamma_k \sum_{s'=1}^{S} w_{ss'} \boldsymbol{x}_{s',t} + \rho \sum_{s'=1}^{S} w_{ss'} y_{s',t-1} + \eta_s \qquad \text{Equation (3)}$$

In Equation (3) above, $x_{s,t}^k$ is the k^{th} observed attribute specific to station s and time-period t, β_k is the corresponding parameter, and K is the total number of observed attributes. The $\sum_{r=1}^{t-1} \alpha_r y_{s,r}$ term captures temporal dynamics whereby the outcomes in previous time-periods affect the outcome in the current time-period t.

The $w_{ss'}$ term captures the spatial weight for the station-pair (ss') and is specified as an inverse function of spatial distance between the two stations. Also, $w_{ss'}$ is normalized to ensure that the spatial influence across all neighboring stations adds up to 1 as follows:

$$\sum_{s'=1}^{S} w_{ss'} = 1 \; and \; w_{ss} = 0 \; \forall \; s \qquad \text{Equation (4)}$$

The $\sum_{k=1}^{K} \gamma_k \sum_{s'=1}^{S} w_{ss'} \boldsymbol{x}_{s',t}$ term captures the spatial spillover effects whereby the outcomes at any given station are affected by the demographic and economic activity in neighboring stations. This spatial spillover effect is captured using spatially weighted explanatory variables as shown in Equation (3) (Anselin, 2003, Anselin, 2013, Bhat et al., 2014). Next, $\rho \sum_{s'=1}^{S} w_{ss'} y_{s',t-1}$captures the spatial dependency between count outcomes in all neighboring stations in the previous time-period 't-1' and the count outcome at station s in time-period t. The α_r, γ_k, and ρ parameters are coefficients corresponding to the temporal, spatial spillover, and spatial lag dependency effects, respectively. Lastly, η_s is the station-specific random effect that influences count outcomes at station s across all time-periods t and is assumed to a stochastic independent and identically distributed (*i.i.d.*) realization from a univariate normal distribution with mean 0 and standard deviation σ.

Assuming $\varepsilon_{s,t,i}$ in Equation (1) to be an *i.i.d.* realization from standard Gumbel distribution across all count outcomes will lead to the MNL model. In this case, the probability of count outcome i conditional on η_s is given by:

$$P(y_{s,t} = i|\eta_s) = \frac{e^{ln\left(\frac{\lambda_{s,t}^i}{i!}\right)+V_{s,t,i}}}{\sum_{j=0}^{\infty} e^{ln\left(\frac{\lambda_{s,t}^j}{j!}\right)+V_{s,t,j}}} \qquad \text{Equation (5)}$$

The probability expression in Equation (5) collapses to the standard Poisson model if $V_{s,t,i} = 0 \; \forall \; i$. To see this, consider the analysis below:

$$P(y_{s,t} = i|\eta_s) = \frac{e^{ln\left(\frac{\lambda_{s,t}^i}{i!}\right)}}{\sum_{j=0}^{\infty} e^{ln\left(\frac{\lambda_{s,t}^j}{j!}\right)}} = \frac{\frac{\lambda_{s,t}^i}{i!}}{\sum_{j=0}^{\infty}\frac{\lambda_{s,t}^j}{j!}} = e^{-\lambda_{s,t}} \times \left(\frac{\lambda_{s,t}^i}{i!}\right) \quad \text{Equation (6)}$$

The Poisson model is not suited for analyzing count data that is typically over-dispersed, *i.e.,* the variance is greater than the mean. So, this study adopted the GEV version of the negative binomial (GEV NB) model where the observed part of $\tilde{V}_{s,t,i}$ is parameterized as follows (Paleti, 2016):

$$\tilde{V}_{s,t,i} = ln\left[\frac{\Gamma(r+i)}{\Gamma(r)\Gamma(i+1)}\left(\frac{\lambda_{s,t}}{1+\lambda_{s,t}}\right)^i\right], \quad \text{Equation (7)}$$

where r is the dispersion parameter and Γ is the Gamma function. This model collapses to the Poisson model when r tends to infinity.

The probability of the vector of count outcomes $\boldsymbol{y}_s = (y_{s,1}, y_{s,2}, \dots . y_{s,T})$ for station s conditional on η_s can be written as:

$$P(\boldsymbol{y}_s|\eta_s) = \prod_{t=1}^{T} \frac{e^{\tilde{V}_{s,t,y_{s,t}} + V_{s,t,y_{s,t}}}}{\sum_{j=0}^{\infty} e^{ln\left(\frac{\lambda_{s,t}^j}{j!}\right) + V_{s,t,j}}} \quad \text{Equation (8)}$$

The summation in the denominator of Equation (8) can be truncated at the maximum observed count value in the data. The unconditional probability of $\boldsymbol{y}_s$ is obtained by integrating out η_s as follows:

$$P(\boldsymbol{y}_s) = \int_{\eta_s=-\infty}^{\eta_s=\infty} \prod_{t=1}^{T} \frac{e^{\tilde{V}_{s,t,y_{s,t}} + V_{s,t,y_{s,t}}}}{\sum_{j=0}^{\infty} e^{ln\left(\frac{\lambda_{s,t}^j}{j!}\right) + V_{s,t,j}}} f(\eta_s) d\eta_s \quad \text{Equation (9)}$$

where $f(\eta_s)$ is the probability density function of univariate normal distribution with mean 0 and standard deviation σ.

4. Data description

The study area adopted for the analysis is the Washington–Arlington–Alexandria region which is the largest metropolitan area in the Census South-East region. The region is home to the Capital Bikeshare system that is one of the oldest and largest public BSSs in the United States. The service launched in September 2010 has over 350 stations and 3,000 bikes across Washington, D.C., Arlington and Alexandria, VA and Montgomery County, MD. The system has logged 2.9 million bike rides with a combined duration of 50 million minutes in 2014. A significant proportion of the BSS data was obtained from the Capital Bikeshare website (https://www.capitalbikeshare.com/system-data). This website provides real-time system state data for each bike station (number of available bikes and empty docks) as well as quarterly trip datasets since 2010. The trip dataset includes information about the origin and destination stations, the start and end times, and the unique bike number of each trip. Data for the year 2016 was extracted from the trip database and individual bike trips were aggregated to compute the number of hourly departures and arrivals for every day of the year for each station. Subsequently, weather information from the National Climatic Data Center, land-use and infrastructure data from Opendata DC (http://opendata.dc.gov/), and demographic information from the 2010 US Census were appended to the BSS demand data. The analysis in this study was undertaken at the spatial resolution of a bike station at a temporal resolution of one hour.

The final variables considered in this study are shown in Table 1. The data corresponding to the time between 11 pm to 6 am where there was very little activity in the system was excluded from the analysis. The remaining active time-window between 6 am and 11 pm accounted for more than 98% of the total activity (both in terms of arrivals and departures) in the system. Time-of-day indicators were defined for every hour between 6 am and 11 pm to account for time-of-day specific average effects on the BSS demand. The demographic data was available at the zip code level. So, all bike stations in the same zip code area have the same demographic variables. Given that the typically preferred walking distance is about 200-400 meters, a 300 meters buffer region was defined for each bike station. All the built environment and infrastructure-related variables were computed within this 300m buffer region. In addition to these explanatory variables, the BSS demand in the previous time-period and spatially weighted demand in the earlier time-period were used to account for temporal and spatial dynamics. The spatial weight matrix was constructed using the quadratic geographic Euclidian distance between stations, *i.e.* the spatial dependency between any two bike stations was assumed to be inversely proportional to the square of the distance between the two stations. Please note that some of the explanatory variables were scaled to ensure smooth convergence during model estimation.

Table 1. Descriptive Statistics

Variable	Description	Min	Max	Mean
Dependent				
Departures	Number of bike departures per hour	0.00	60.00	1.09
Arrivals	Number of bike arrivals per hour	0.00	60.00	1.19
Temporal				
6am to 11pm	Time-of-day indicator	0.00	1.00	0.06
Weekend	1 if the prediction day is a weekend/ otherwise 0	0.00	1.00	0.23
Tue to Sun	Day-of-week indicator	0.00	1.00	-
Holiday	1 if prediction is a Holiday/ otherwise 0	0.00	1.00	-
Weather				
Temp.	Temperature (F)	12.90	100	63.36
Temp. above 75	1 if temperature is above 75 F/ otherwise 0	0.00	1.00	0.32
Percip.	Average Hourly Precipitation (in)	0.00	0.37	0.01
Rainy	1 if it is snowy or rainy/ otherwise 0	0.00	1.00	0.17
Socio-Demographics				
Pop. Density	Population density (10,000 people per square mile)	0.00	31.63	0.62
Female Proportion	The ratio of women older than 16 years and total	0.46	0.59	0.52

Variable	Description	Min	Max	Mean
	population			
Workers Proportion	The ratio of workers above 16 years and total population	0.26	0.83	0.68
Income <50k	Number of households with income less than $50K/10000	0.00	0.71	0.26
Income <50k & >100k	Number of households with income $50k to $100k/10000	0.00	0.38	0.24
Income <100k & >200k	Number of households with income $100k to $200k/10000	0.00	1.00	0.31
Income >200k	Number of households with income more than $200k/10000	0.00	0.50	0.16
Built Environment				
# of Hotels	Number of hotels in the buffer region/10	0.00	0.60	0.03
# of Museums	Number of museums in the buffer region	0.00	2.00	0.07
# of Rec. Center	Number of recreation centers in the buffer region	0.00	3.00	0.07
# of Bussiness Center	Number of business centers in the buffer region/100	0.00	55.00	2.06
# of Universities	Number of universities in the buffer region	0.00	1.00	0.02
# of Schools	Number of schools in the buffer region/10	0.00	0.30	0.01
# of Green Spaces	Number of green spaces in the buffer region/100	0.00	1.58	0.06
MD	1 if station is in Maryland/ otherwise 0	0.00	1.00	0.14
VA	1 if station is in Virginia/ otherwise 0	0.00	1.00	0.31
CBD	1 if station is in Central Business District/ otherwise 0	0.00	1.00	0.03
Infrastructure				
# of Bike Stations	Number of bike stations in the buffer region/10	0.10	0.20	0.11
# of Car Sharing	Number of car sharing units in the buffer region/10	0.00	0.50	0.03
# of Bus Stops	Number of bus stops in the buffer region/100	0.00	0.10	0.02
# of Metro Stations	Number of metro stations in the buffer region	0.00	1.00	0.10
# of Train Stations	Number of train stations in the buffer region/10	0.00	1.50	0.02
# of Shuttle Bus Stops	Number of shuttle bus stops in the buffer region/10	0.00	7.00	0.41
# of Tourist Bus Stops	Number of tourist bus stops in the buffer region	0.00	2.00	0.04
Temporal Dynamics				
Prev. Hour departures	Previous hour departures	0.00	60.00	1.09
Prev. Hour arrivals	Previous hour arrivals	0.00	60.00	1.19
Prev. Week Departure	Previous week, same day-of-week, same hour departures	0.00	60.00	1.09
Prev. Week Arrival	Previous week, same day-of-week, same hour arrivals	0.00	60.00	1.19
Spatial Dynamics				
Prev. Hour Weighted Departures	Spatially weighted departures in the previous hour	0.04	8.96	1.84
Prev. Hour Weighted Arrivals	Spatially weighted arrivals in the previous hour	0.03	8.93	1.86

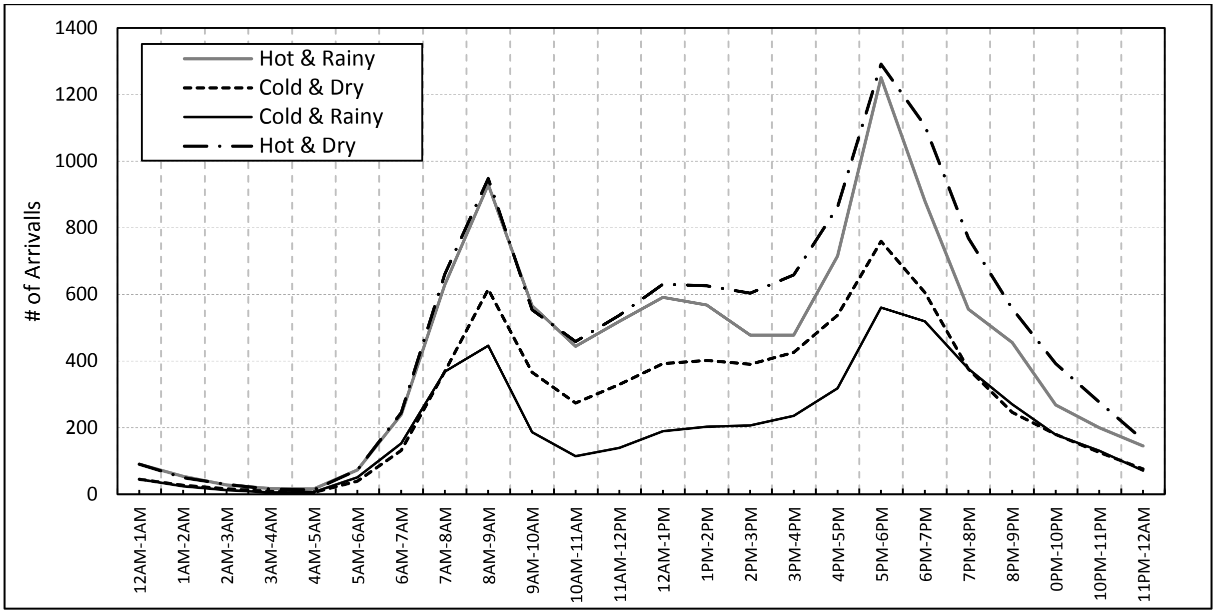

Figure 1. Average Hourly Arrivals across All Stations within Each Cluster

Table 2. Weather Clustering Analysis Results

Cluster	Size (# of Days)	Temperature (º F)			Precipitation (in)			Label
		Mean	Max.	Min.	Mean	Max.	Min.	
1	9	77.52	98.10	54.00	0.01	0.93	0.00	Hot and Rainy
2	170	44.70	73.90	12.90	0.00	0.21	0.00	Cold and Dry
3	13	52.08	68.00	30.90	0.01	0.37	0.00	Cold and Rainy
4	174	74.52	100.00	45.00	0.00	0.31	0.00	Hot and Dry

The BSS demand varies considerably as a function of weather conditions. So, clustering analysis was undertaken to categorize days based on the daily average temperature and precipitation recordings using the K-Means clustering method. The temperature and precipitation readings were normalized during the clustering analysis to ensure that the temperature attribute does not completely dominate the clustering algorithm due to its higher scale. Based on the clustering analysis, the 366 days of 2016 were grouped into four clusters that are labeled as 'hot and wet, 'cold and dry', 'cold and wet', and 'hot and dry' days. Table 2 presents the description of the four clusters and Figure 1 depicts the average total arrivals across all stations and days in the four clusters. It can be seen from Figure 1 that the demand levels vary considerably across the four clusters.

5. Model Results

The model estimation process was guided by past research findings, parameter intuitiveness, and statistical significance considerations. Only parameters that were significant at the 95% confidence level were retained. The results of the best specification of the GEV NB count are presented in the paper. Dataset is divided into two subsets: training and testing. For estimation purposes, the training dataset is employed including 300 stations, 10 days, and 5 time-periods per day that were randomly sampled from the complete sample within each cluster, i.e. $S = 300$ and $T = 10 \times 5 = 50$. For cluster 1, however, all nine days were used in the estimation analysis. Separate models were developed for departures and arrivals for each cluster leading to a total of eight different models. Tables 3 and 4 present the results of the expected count parameter component $\lambda_{s,t}$ for arrivals and departures, respectively.

Table 3. Expected Count Component of GEV NB Models for Arrivals

Variables	Hot and Rainy		Cold and Dry		Cold and Rainy		Hot and Dry	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
'CONSTANT'	-0.52	-0.58	1.81	6.95	1.46	4.35	2.20	10.59
7am to 8am	0.76	12.33	0.78	11.20	0.83	11.72	0.86	15.11
8am to 9am	0.71	12.07	0.69	10.45	0.74	10.89	0.79	15.51
9am to 10am							0.16	2.73
10am to 11am	0.35	5.83	0.46	6.61	0.48	6.62	0.43	7.80
11am to 12pm	0.39	6.60	0.45	6.59	0.46	6.54	0.53	10.17
12pm to 1pm	0.23	3.90	0.26	3.88			0.26	4.89
1pm to 2pm							0.22	4.06
2pm to 3pm	0.17	2.65	0.25	3.66	0.25	3.58	0.45	8.67
3pm to 4pm	0.68	11.41	0.49	7.41	0.52	7.60	0.56	10.97
4pm to 5pm	0.99	17.75	0.58	9.19	0.58	8.91	0.80	16.38
5pm to 6pm			-0.31	-4.07	-0.28	-3.57		
6pm to 7pm			-0.38	-4.86	-0.39	-4.81		
8pm to 9pm			-0.36	-4.57				
9pm to 10pm	-0.33	-4.70	-0.27	-3.28	-0.28	-3.29		
10pm to 11pm	-0.36	-4.37	-0.56	-6.07	-0.60	-6.44	-0.19	-2.94
Rainy	-0.50	-11.55	-0.48	-5.16			-0.40	-4.74
Female Proportion	-5.99	-11.83	-5.87	-11.59	-6.78	-12.13	-5.29	-13.10
Workers Proportion					1.04	6.22		
Income >50k	0.48	6.82						
# of Hotels	0.88	4.11	1.40	5.75	1.42	5.44	0.95	5.58
# of Car Sharing			0.85	4.48				
# of Metro Stations	0.21	4.97	0.17	3.87	0.15	3.09		
# of Universities					0.33	3.15	0.17	2.64
# of Green Spaces							0.15	2.12
# of Tourist Bus Stops			0.23	3.75			0.18	3.88
Pop. Density	-0.01	-2.23	-0.02	-2.58			-0.02	-2.53
MD	-1.34	-15.98	-1.15	-15.78	-1.26	-16.93	-1.10	-17.72
VA	-0.81	-17.79	-0.86	-18.09	-1.11	-22.25	-0.72	-19.38
Prev. Hour arrivals	0.08	7.70	0.13	12.73	0.15	14.08	0.08	14.25
Prev. Hour departures	0.11	9.52	0.07	7.82	0.08	7.84	0.02	2.58
Prev. Hour Weighted Arrivals	0.18	11.21	0.37	17.77	0.37	16.64	0.16	13.22

Table 4. Expected Count Component of GEV NB Models for Departures

Variables	Hot and Rainy		Cold and Dry		Cold and Rainy		Hot and Dry	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
'CONSTANT'	-0.43	-7.12	0.40	1.79	-1.01	-0.97	0.26	0.40
7am to 8am	0.64	12.88	1.03	12.86	0.80	8.80	0.72	11.30
8am to 9am	0.31	6.18	0.96	12.53	0.67	7.39	0.51	8.43

10am to 11am	-0.29	-4.89			-0.28	-2.76	-0.31	-5.16
11am to 12pm			0.38	4.19				
12pm to 1pm	0.32	6.33	0.54	5.83	0.19	1.83	0.41	5.60
1pm to 2pm	0.36	7.08	0.66	6.72	0.65	6.40	0.50	7.15
2pm to 3pm	0.31	6.13	0.50	4.92	0.31	2.99	0.36	4.88
3pm to 4pm			0.33	3.64			0.19	2.68
4pm to 5pm	0.18	3.30	0.50	5.41	0.18	2.08	0.24	3.34
5pm to 6pm	0.65	12.99	0.82	10.15	0.63	7.40	0.74	9.99
6pm to 7pm	0.96	21.82	1.03	10.66	1.28	11.61	0.92	13.38
7pm to 8pm			0.30	3.29	0.41	5.14	0.25	3.91
8pm to 9pm	0.17	3.17						
9pm to 10pm					-0.25	-2.65		
10pm to 11pm					-0.43	-4.08	-0.27	-4.11
Weekend			-0.20	-5.05				
Temp.			0.01	4.62				
Rainy	-0.48	-12.93	-0.53	-4.45	-0.29	-2.46	-0.42	-4.02
Female Proportion			-6.70	-19.71	-7.08	-18.52	-5.41	-21.41
Workers Proportion			1.43	15.16	1.61	15.62		
Income >50k	0.25	4.57					0.58	18.82
# of Hotels			2.39	15.00	2.01	14.16	1.75	14.63
# of Metro Stations	0.17	4.98	0.20	6.98	0.18	5.81	0.20	10.58
# of Shuttle Bus Stops	0.04	4.32						
Pop. Density	-0.03	-3.20	-0.03	-3.89	-0.03	-6.23	-0.04	-11.17
MD	-1.20	-16.74	-1.56	-27.36	-1.48	-23.19	-1.37	-27.07
VA	-0.62	-16.86	-1.10	-31.67	-1.46	-31.38	-0.85	-28.45
Prev. Hour arrivals	0.11	20.81	0.13	12.00	0.26	9.78	0.11	10.57
Prev. Week Arrival	0.04	10.96			0.08	9.61	0.05	8.22
Prev. Hour departures			0.09	7.37				
Prev. Hour Weighted Arrivals	0.18	13.58	0.24	7.77	0.43	14.16	0.18	11.92

Several hour-of-day indicator variables were statistically significant indicating time-of-day specific average effects for all clusters. Although the demand patterns tend to vary significantly between weekends and weekdays, the weekend indicator variable was significant only for the cluster 2 arrivals model. This is probably because of the temporal and spatial dynamics effects that were already captured in the model which negated the need for the weekend indicator variable. The effects of all the covariates in the model were segmented by the four clusters defined based on weather conditions. However, it is still possible that the temperature and precipitation variation within each cluster will lead to changes in the arrival and departure patterns. To capture these effects within each cluster, temperature and precipitation variables were tested in the model. As expected, warmer temperatures and lower precipitation levels were associated with higher departures and arrivals although these effects were not significant in all clusters. A higher proportion of workers and a lower proportion of women in the zip code area containing the bike station were associated with higher demand levels. Higher income (>$50k) population composition was associated with more departures and arrivals. Interestingly, higher population density levels were associated with lower BSS demand levels. This is probably because higher population density levels indicate residential areas with limited bike infrastructure and fewer activity-opportunity centers. A positive correlation between bike sharing system infrastructure and cycling demand is shown in the literature (Yanocha et al., 2018). Green spaces, universities, and hotels that are activity-opportunity centers were found to increase both bike arrivals and departures. As expected, better accessibility to other modes of transportation was associated with higher BSS demand. Specifically, the number of car sharing stations, regional metro stations, commute bus stops, and tourist bus stops within the 300m buffer around each station were positively correlated with the BSS demand. These results demonstrate the utility of public BSSs to solve the first/last mile problem of transit users (Nair et al., 2013, Cohen and Shaheen, 2018). Results corresponding to indicator variables for the area of service indicate, on average, higher demand levels in Washington DC compared to Arlington and Alexandria, VA and Montgomery County, MD. Lastly, the demand levels in a certain hour-of-day were positively associated with higher demand levels in the previous hour, higher demand levels in the same hour-of-day on the same day-of-week in the earlier week, and higher demand levels in all neighboring stations. These results confirm the presence of strong temporal and spatial dynamic effects in the BSS arrival and departures patterns.

Table 5 presents the results corresponding to the additional propensity component $\tilde{V}_{s,t,i}$ in all the models. Several parameters were significant in the $V_{s,t,i}$ component for count outcomes zero to six indicating that there are considerable deviations in the probability mass implied by the standard NB model for these specific count outcomes. In addition to several count outcome specific average effects (*i.e.,* the constants in Table 5), higher demand levels in the previous hour and higher demand levels in the previous week on the same day-of-week and same hour were associated with lower probability mass for lower count outcomes. Also, as expected, rainy weather was associated with a higher probability mass for lower count outcomes. These results underscore the importance of the additional flexibility in the GEV count modeling framework compared to standard count models.

Table 5. Additional Propensity Component in GEV NB Models

	Count	Cluster	Rainy	PHDep	PWDep	PHArr	WNDep	Constant
Departures	Zero	1		-0.157	0.125	-0.220		-4.743
		2		-0.416				
		3	0.816	-0.392	-0.051			
		4		-0.399		-0.192		0.783
	One	1			0.137	-0.046		-2.908
		3	-0.007					
		4		-0.187				
	Two	1			0.107	-0.046		-1.957
		3	0.361					
		4		-0.125				
	Three	1			0.074			-1.462
		4		-0.093				
	Four	1			0.062			-0.936
		4		-0.045				
	Five	1			0.052			-0.579
		4		-0.045				
	Six	1			0.052			-0.579
Arrivals	Zero	1		-0.310	-0.092		-0.111	1.390
		2		-0.436	-0.064		-0.265	
		3	1.023	-0.319				-3.298

	4		-0.410	-0.085		-0.131	-2.134
One	2		-0.186				
	3	0.706					-1.230
	4		-0.136				-1.370
Two	3	0.524					-0.581
	4		-0.099				-0.848
Three	4		-0.047				-0.580
Four	4		-0.045				-0.308
Five	4						-0.268

6. Performance Evaluation

For validation purposes, the testing dataset is employed including 300 stations, 5 days, and 5 time-periods per day that were randomly sampled within each cluster after excluding the training subset, i.e. $\mathrm{S} = 300$ and $\mathrm{T} = 5 \times 5 = 25$. The predictive power of the proposed models was evaluated using several evaluation metrics. First, the Proportion of Total Demand (PTD) metric was computed as the ratio of the predicted and observed demand across all stations and all days in the entire system. Second, the Mean Absolute Error (MAE) of prediction was computed as the average value of the absolute difference between observed and predicted demand in each hour across all days and stations. Third, the Root Mean Squared Error (RMSE) was computed as the square root of the average sum of squared errors in the predictions. Fourth, the Percentage Predicted Correctly (PPC) metric was calculated as the percentage of cases when the predicted demand is exactly equal to the observed demand (e.g., observed demand at a station is 2 arrivals and the predicted demand is also 2 arrivals). Lastly, the distribution of the absolute prediction error was also reported. For example, an absolute error of one implies that the predicted and observed counts differ by ± 1. Table 6 presents the results of the prediction analysis. Both the arrivals and departures models were able to predict the total demand across all stations within 1% and 4% error, respectively although these percentages can go up to 8% in some of the clusters. The MAE values were 1.19 and 1.23 and the corresponding RMSE values were 2.63 and 2.83 for arrivals and departures, respectively. These numbers indicate that the margin of error in the predicted demand is about 1 in both arrivals and departures. In fact, more than 50% of predictions exactly match the observed demand and the margin of error is within one for about 75% of the predictions as indicated by the PPC and absolute errors in Table 6. In order to evaluate the proposed model, the model performance was compared with the commonly used negative binomial regression (referred to as the “Base Model”). It can be seen from Table 6 that the proposed model has significantly better prediction power over the base models indicated by relatively lower MAE and RMSE values for arrivals and departures.

Table 6. Performance Metrics

Demand	Performance Metric	Proposed Model					Base Model
		Cluster					
		Hot & Rainy	Cold & Dry	Cold & Rainy	Hot & Dry	Average	
Arrivals	Proportion of Total Demand (PTD)	1.02	1.03	1.01	0.99	1.01	1.59
	Mean Absolute Error (MAE)	1.38	1.05	0.76	1.57	1.19	2.09

	Root Mean Squared Error (RMSE)	2.92	2.46	1.95	3.21	2.63	4.21
	Percentage Predicted Correctly (PPC)	49.88	58.45	64.65	45.43	54.60	41.52
	Absolute Error Distribution (%)						
	Zero	49.88	58.45	64.65	45.43	54.60	41.52
	One	22.59	21.39	19.99	23.41	21.85	24.83
	Two	10.62	8.47	7.13	11.47	9.42	13.36
	Three	5.93	4.06	3.60	6.79	5.09	9.06
	Four or more	10.98	7.63	4.63	12.90	9.03	11.23
Departures	Proportion of Total Demand (PTD)	1.01	1.08	1.08	1.01	1.04	1.72
	Mean Absolute Error (MAE)	1.43	1.08	0.81	1.60	1.23	2.11
	Root Mean Squared Error (RMSE)	2.95	2.86	2.11	3.38	2.83	5.08
	Percentage Predicted Correctly (PPC)	48.66	57.56	63.89	45.78	53.97	39.98
	Absolute Error Distribution (%)						
	Zero	48.66	57.56	63.89	45.78	53.97	39.98
	One	22.33	22.37	20.53	23.60	22.21	25.25
	Two	10.81	8.39	7.21	11.19	9.40	14.03
	Three	6.47	4.33	3.04	6.24	5.02	7.28
	Four or more	11.73	7.34	5.32	13.19	9.39	13.46

7. Research Applicability

Currently, there are multiple barriers in the path to the adoption of the advanced statistical models developed for predicting BSS demand. These include both technical (e.g., complex models that need significant computational resources) and institutional (e.g., budgetary constraints) barriers. The current study aims to develop an *accurate yet tractable* modeling method that can predict bike-sharing demand in real-time, thereby reducing the technical complexity barrier associated with most models available in the literature. However, the complex spatio-temporal patterns and non-standard distributional properties associated with bike sharing demand data present unique modeling challenges. The statistical model developed in this study accomplishes the research objectives by ensuring a closed-form probability expression for all possible outcomes. The model does this by re-casting the count model as a special case of the well-known MNL model, implying that forecasting using the models developed in this paper has the level of complexity as the standard MNL model. So, the entire forecasting system can be made operational in a simple excel-based tool because the model probabilities can be translated into predicted outcomes using simple Monte-Carlo method or rule-based methods (ex., alternative with the highest probability is the realized outcome). So, bike sharing agencies can easily implement the model in their preferred software platform (e.g., excel) given its close semblance to the well-known MNL model. While model applicability was demonstrated for the Capital Bikeshare system, the model is generalizable to any bike-sharing system with arbitrary distributional properties for the demand data. The model parameters may need adjustment based on new training data when applied to bike-sharing systems other than Capital Bikeshare, but the basic structure and prediction mechanism are generalizable.

8. Summary and Conclusion

Real-time prediction of public bike sharing systems (BSSs) is needed for optimizing the day-to-day operations of the system (e.g., optimal rebalancing plans based on expected future demand patterns) and developing advisory tools for customers (e.g., a smartphone mobile app that indicates the expected bike and dock availability at bike stations). In this context, the current study developed Generalized Extreme Value (GEV) count models to predict the hourly number of arrivals and departures at each bike station. The model developed encompasses methodological components that (1) provide additional flexibility to account for over or under-representation of multiple count outcomes, (2) account for station-specific unobserved effects that influence demand on all days for a given station, and (3) account for spatial and temporal dynamics. The proposed model was used to analyze the demand patterns of the Capital Bikeshare system that serves Washington, D.C., Arlington and Alexandria, VA and Montgomery County, MD. Clustering methods were used to group the 366 days in 2016 into four clusters based on average daily temperature and precipitation values. Next, separate models were developed for hourly arrivals and departures. The models uncovered significant time-of-day, weather, built environment, infrastructure, and temporal and spatial dependence effects. The predictive power of the proposed models was evaluated using several metrics. Overall, the results indicate reasonably good predictive performance of the proposed models both at the aggregate system and disaggregate station and time-of-day levels.

There are several possible avenues for future research. First, the expected demand patterns predicted by the models developed in this study can serve as input to an optimization module with quick heuristic algorithms to determine optimal rebalancing routes and schedules for a given fleet size and locations of the rebalancing trucks. Second, the usage levels of the BSS are closely tied to the quality of transit infrastructure in the region. In some cases, people use the BSS as an access or egress mode for transit whereas in other cases people stop using transit and instead use the BSS for undertaking a trip. There has been very little concerted effort to explore the synergy between traditional and innovative public transportation investments. The real-time system state prediction models developed in this study would serve as a good starting point for studying nature (substitution or complementarity) and the extent of interdependencies between BSS and transit. Specifically, a systems approach to modeling multi-modal interactions will help identify the locations and the extent to which transit and BSS complement and compete. So, low-cost options such as optimal service configurations can be pursued for enhancing transit ridership and BSS usage levels. Third, exploring the transferability of models developed in this study for other BSS systems in the nation is an interesting research avenue.

Acknowledgment

This study was funded by the Mid-Atlantic Transportation Sustainability Region 3 University Transportation Center (MATS UTC).

REFERENCES

AHMED, F., ROSE, G. & JACOB, C. Impact of weather on commuter cyclist behaviour and implications for climate change adaptation. Australasian Transport Research Forum (ATRF), 2010 Canberra, Australia.

ANSELIN, L. 2003. Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review,* 26**,** 153-166.

ANSELIN, L. 2013. *Spatial econometrics: methods and models*, Springer Science & Business Media.

BHAT, C. R., PALETI, R. & SINGH, P. 2014. A spatial multivariate count model for firm location decisions. *Journal of Regional Science,* 54**,** 462-502.

BÖCKER, L. & THORSSON, S. 2014. Integrated weather effects on cycling shares, frequencies, and durations in Rotterdam, the Netherlands. *Weather, Climate, and Society,* 6**,** 468-481.

COHEN, A. & SHAHEEN, S. 2018. *Planning for shared mobility*. American Planning Association. 1611901863. Retrieved from: <https://escholarship.org/content/qt0dk3h89p/qt0dk3h89p.pdf> (July 2019).

CONTARDO, C., MORENCY, C. & ROUSSEAU, L.-M. 2012. *Balancing a dynamic public bike-sharing system*, Cirrelt Montreal.

DELL'AMICO, M., IORI, M., NOVELLANI, S. & SUBRAMANIAN, A. 2018. The bike sharing rebalancing problem with stochastic demands. *Transportation research part B: methodological,* 118**,** 362-380.

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.

ERMAGUN, A., LINDSEY, G. & LOH, T. H. 2017. Urban trails and demand response to weather variations. *Transportation Research Part D: Transport and Environment,* 63**,** 404-420.

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.

FAGHIH-IMANI, A., ELURU, N. & PALETI, R. 2017. How bicycling sharing system usage is affected by land use and urban form: analysis from system and user perspectives. *European Journal of Transport and Infrastructure Research,* 17**,** 425-441.

FLYNN, B. S., DANA, G. S., SEARS, J. & AULTMAN-HALL, L. 2012. Weather factor impacts on commuting to work by bicycle. *Preventive Medicine,* 54**,** 122-124.

FROEHLICH, J., NEUMANN, J. & OLIVER, N. Sensing and predicting the pulse of the city through shared bicycling. International Joint Conferences on Artificial Intelligence Organization, 2009 Pasadena, California. AAAI, 1420-1426.

GALLOP, C., TSE, C. & ZHAO, J. 2011. A seasonal autoregressive model of Vancouver bicycle traffic using weather variables. *i-Manager's Journal on Civil Engineering,* 1**,** 9.

GHOSH, S., VARAKANTHAM, P., ADULYASAK, Y. & JAILLET, P. 2017. Dynamic Repositioning to Reduce Lost Demand in Bike Sharing Systems. *Journal of Artificial Intelligence Research,* 58**,** 387-430.

GOTSCHI, T. & MILLS, K. 2008. *Active transportation for America: The case for increased federal investment in bicycling and walking*. Rails to Trails Conservancy. Retrieved from: <https://trid.trb.org/view/874547> (July 2019).

GREENE, W. 2008. Functional forms for the negative binomial model for count data. *Economics Letters,* 99**,** 585-590.

GURMU, S. 1998. Generalized hurdle count data regression models. *Economics Letters,* 58**,** 263-268.
HANDY, S. L., XING, Y. & BUEHLER, T. J. 2010. Factors associated with bicycle ownership and use: a study of six small US cities. *Transportation,* 37**,** 967-985.
JONES, M. G., RYAN, S., DONLON, J., LEDBETTER, L., RAGLAND, D. R. & ARNOLD, L. S. 2010. *Seamless travel: Measuring bicycle and pedestrian activity in San Diego County and its relationship to land use, transportation, safety, and facility type*. Institute of Transportation Studies, University of California, Berkeley. 1055-1425. Retrieved from: <https://safetrec.berkeley.edu/sites/default/files/publications/seamless_travel.pdf> (July 2019).
KALTENBRUNNER, A., MEZA, R., GRIVOLLA, J., CODINA, J. & BANCHS, R. 2010. Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing,* 6**,** 455-466.
LEWIN, A. Temporal and weather impacts on bicycle volumes. Transportation Research Board 90th Annual Meeting, 2011 Washington DC, United States. Transportation Research Board
LI, Y., ZHENG, Y., ZHANG, H. & CHEN, L. Traffic prediction in a bike-sharing system. Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2015 Seattle, Washington. ACM, 33.
LINDSEY, G., WILSON, J., ANNE YANG, J. & ALEXA, C. 2008. Urban greenways, trail characteristics and trail use: implications for design. *Journal of Urban Design,* 13**,** 53-79.
LINDSEY, G., WILSON, J., RUBCHINSKAYA, E., YANG, J. & HAN, Y. 2007. Estimating urban trail traffic: Methods for existing and proposed trails. *Landscape and Urban Planning,* 81**,** 299-315.
LIU, J., SUN, L., CHEN, W. & XIONG, H. Rebalancing bike sharing systems: A multi-source data smart optimization. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016 San Francisco, California. ACM, 1005-1014.
LORD, D., WASHINGTON, S. P. & IVAN, J. N. 2005. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis & Prevention,* 37**,** 35-46.
MIRANDA-MORENO, L. & NOSAL, T. 2011. Weather or not to cycle: Temporal trends and impact of weather on cycling in an urban environment. *Transportation Research Record,* 2247**,** 42-52.
MOUDON, A. V., LEE, C., CHEADLE, A. D., COLLIER, C. W., JOHNSON, D., SCHMID, T. L. & WEATHER, R. D. 2005. Cycling and the built environment, a US perspective. *Transportation Research Part D: Transport and Environment,* 10**,** 245-261.
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**,** 85-106.
NARAYANAMOORTHY, S., PALETI, R. & BHAT, C. R. 2013. On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. *Transportation research part B: methodological,* 55**,** 245-264.
PALETI, R. 2016. Generalized Extreme Value models for count data: Application to worker telecommuting frequency choices. *Transportation Research Part B: Methodological,* 83**,** 104-120.

PALETI, R., CETIN, M., SOHRABI, S., BALAN, L., RAKHA, H. & ELHENAWY, M. 2018. *Real-time system state prediction & optimal rebalancing strategies for public bike sharing systems*. Mid-Atlantic Transportation Sustainability Center – Region 3 University Transportation Center. Retrieved from: <http://www.matsutc.org/wp-content/uploads/2018/02/BikeSharingProject_FinalReport.pdf> (July 2019).

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.

PHUNG, J. & ROSE, G. Temporal variations in usage of Melbourne's bike paths. Proceedings of 30th Australasian Transport Research Forum, 2007 Melbourne. ATRF.

PIATKOWSKI, D. P., KRIZEK, K. J. & HANDY, S. L. 2015. Accounting for the short term substitution effects of walking and cycling in sustainable transportation. *Travel Behaviour and Society,* 2**,** 32-41.

RIXEY, R. 2013. Station-level forecasting of bikesharing ridership: Station Network Effects in Three US Systems. *Transportation Research Record,* 2387**,** 46-55.

SCHNEIDER, R., ARNOLD, L. & RAGLAND, D. 2009. Pilot model for estimating pedestrian intersection crossing volumes. *Transportation Research Record,* 2140**,** 13-26.

SCHNEIDER, R. J., HENRY, T., MITMAN, M. F., STONEHILL, L. & KOEHLER, J. 2012. Development and application of the San Francisco pedestrian intersection volume model. *Transportation Research Record,* 2299**,** 65-78.

SCHUIJBROEK, J., HAMPSHIRE, R. C. & VAN HOEVE, W.-J. 2017. Inventory rebalancing and vehicle routing in bike sharing systems. *European Journal of Operational Research,* 257**,** 992-1004.

SHAHEEN, S., GUZMAN, S. & ZHANG, H. 2010. Bikesharing in Europe, the Americas, and Asia: Past, present, and future. *Transportation Research Record,* 2143**,** 159-167.

SHAHEEN, S. A., COHEN, A. P. & MARTIN, E. W. 2013. Public bikesharing in North America: early operator understanding and emerging trends. *Transportation Research Record,* 2387**,** 83-92.

WANG, J., HANKEY, S., WU, X. & LINDSEY, G. 2016. Monitoring and modeling of urban trail traffic: Validation of direct demand models in Minneapolis, Minnesota, and Columbus, Ohio. *Transportation Research Record,* 2593**,** 47-59.

WANG, X., LINDSEY, G., HANKEY, S. & HOFF, K. 2013. Estimating mixed-mode urban trail traffic using negative binomial regression models. *Journal of Urban Planning and Development,* 140**,** 04013006.

WINTERS, M., DAVIDSON, G., KAO, D. & TESCHKE, K. 2011. Motivators and deterrents of bicycling: comparing influences on decisions to ride. *Transportation,* 38**,** 153-168.

YANOCHA, D., MASON, J., PATLAN, M., BENICCHIO, T., ALFRED, I. & LAKSMANA, U. 2018. *The Bikeshare Planning Guide*. Institute for Transportation and Development Policy, New York. Retrieved from: <https://itdpdotorg.wpengine.com/wpcontent/uploads/2013/12/BSPG_digital.pdf> (July 2019).

YOON, J. W., PINELLI, F. & CALABRESE, F. Cityride: a predictive bike sharing journey advisor. IEEE 13th International Conference on Mobile Data Management (MDM), 2012 Bengaluru, India. IEEE, 306-311.