

A bikeshare station area typology to forecast the station-level ridership of system expansion

Steven R. Gehrke
Metropolitan Area Planning Council
sgehrke@mapc.org

Timothy F. Welch
Georgia Institute of Technology
tim.welch@design.gatech.edu

Abstract: The continuous introduction and expansion of docked bikeshare systems with publicly available origin-destination data have opened exciting avenues for bikeshare research. In response, a flux of recent studies has examined the sociodemographic determinants and safety or natural environment deterrents of system ridership. An increasing abundance of disaggregate spatial data has also spurred recent calls for research aimed at extending the utility of these contextual data to model bikeshare demand and trip patterns. As planners and operators seek to expand bikeshare services into underserved areas, a need exists to provide a data-driven understanding of the spatial dynamics of bikeshare use. This study of the Washington, DC, metro region's Capital Bikeshare (CaBi) program answers this call by performing a latent class cluster analysis to identify five bikeshare station area types based on variation in a set of land development pattern, urban design, and transportation infrastructure features. This typology is integrated into a planning application exploring the potential for system expansion into nearby jurisdictions and forecasting the associated trip-making potential between existing and proposed station locations.

Article history:

Received: March 13, 2018
Received in revised form:
November 6, 2018
Accepted: January 10, 2019
Available online: April 12, 2019

1 Introduction

As anticipated, a virtuous cycle of bikeshare system introduction and expansion has taken hold in American cities over the past decade (DeMaio, 2009). The potential to alleviate environmental concerns related to traffic congestion and greenhouse gas emissions as well as a prospect to improve individual mobility and health-related outcomes has likely helped to propel this collective adoption (Shaheen, Guzman, & Zhang, 2010). These docked bikeshare programs are characterized by their provision of bikes at self-serving docking stations that support a quick and easy rental of bikes by riders via advanced user interface technologies (Shaheen et al., 2010; Fishman, 2016).

Launched in fall 2010, the Capital Bikeshare (CaBi) system in Washington, DC was one of these earliest third-generation bikeshare programs to be introduced in the United States. During its first quarter of operation, CaBi riders performed 117,692 trips between the system's 107 stations. However, by the final three months of 2016, the CaBi system had expanded to include 433 bikeshare stations, which supported 771,068 bicycling trips. This system-wide expansion included the development of stations in several neighboring jurisdictions; thus, enabling system riders to now travel to and from a greater diversity of activity locations situated in a variety of environmental contexts.

Copyright 2019 Steven R. Gehrke and Timothy F. Welch

<http://dx.doi.org/10.5198/jtl.2019.1395>

ISSN: 1938-7849 | Licensed under the [Creative Commons Attribution – Noncommercial License 4.0](https://creativecommons.org/licenses/by-nc/4.0/)

CaBi's success has not been an isolated phenomenon, as the number of cities offering bikeshare has increased rapidly in recent years (Fishman, 2016). A dramatic rise in bikeshare-related research has accompanied this nationwide escalation in system introduction and ridership. To date, these studies have focused on understanding the sociodemographic or built environment determinants and safety or natural environment deterrents of bikeshare system ridership (Fishman, 2016). Moreover, a growing abundance of disaggregate spatial bikeshare data has spurred the assertion of a broad research agenda including the described need to extend the utility of contextual data to model bikeshare demand and trip patterns (Corcoran & Li, 2014). Specifically, the spatial configuration and complex interaction of various built environment features related to activity intensity and location, street network design, and transit connectivity complicate the ability to accurately evaluate the travel behaviors and patterns of bikeshare users (Ma, Liu, & Erdogan, 2015).

In response, this study assesses the surrounding spatial contexts of CaBi station areas toward proposing a bikeshare station area typology. Classifying bikeshare station areas based on variation across several important built environment features has the potential to offer an objective, data-driven approach to siting docked bikeshare stations and permit a deeper understanding of the spatial dynamics of bikeshare activity (Corcoran et al., 2014). In a demonstrated planning application, the second part of this study utilizes this typology to forecast the potential ridership associated with a simulated expansion of the CaBi system. The results from this introduction of a bikeshare station area typology using a latent class clustering analysis and subsequent real-world demonstration within a station location planning application are then briefly discussed.

2 Literature review

The introduction of bikeshare systems and expansion of current programs with widely available origin-destination data has opened new avenues for bikeshare research on travel behaviors and patterns (O'Brien, Cheshire, & Batty, 2014). Accurate ridership estimates and travel patterns accompanied by disaggregate built environment information describing station areas have resulted in a number of recent studies exploring the built environment effects on ridership and station-level forecasts (Rixey, 2013). However, since bikeshare is a relatively recent concept in the United States, the evidence base has thus far offered limited details regarding the spatial determinants of its adoption or physical characteristics of the systems themselves (Bachand-Marleau, Lee, & El-Geneidy, 2012). Accordingly, there has been insufficient empirical guidance on the best built environment metrics to describe bikeshare usage or station area distinction (Buck & Buehler, 2012).

Faghih-Imani, Eluru, El-Geneidy, Rabbat, and Haq (2014) explored a suite of factors impacting both bikeshare flows and ridership for Montreal's system. In terms of land development patterns, the authors found that bikeshare stations located in zones with a high population density had increased arrival and departure rates; whereas, only stations with high employment density had a positive association with arrival rates in the morning peak period. A station-level study of bikeshare use in Toronto revealed that trip generation and attraction were both positively influenced by population and employment density on the weekdays, but only population density on the weekends (El-Assi, Mahmoud & Habib, 2017). Studies of the CaBi program have echoed the impact of station area population density on use by noting its positive connection to daily rides (Buck and Buehler, 2012) and monthly rentals (Rixey, 2013).

The latter study also indicated greater retail employment in a bikeshare station area was associated with increased monthly rentals across a pooled sample of three American programs (Rixey, 2013). Meanwhile, the importance of retail services and employment opportunities to bikeshare station placement was previously identified in a market study of bikeshare station and rider demand in Philadelphia

(Krykewycz, Puchalsky, Rocks, Bonnette, & Jaskiewicz, 2010). In an investigation of the attraction of bikeshare to various activity destinations, a study of the Chicago bikeshare program found that an increased number of restaurants within a 300-meter station area buffer had a positive association with ridership (Faghih-Imani & Eluru, 2014). Similarly, Wang, Lindsey, Schoner, and Harrison (2016) modeled a positive link between food-related businesses located within a 200-meter areal buffer and station use. Acknowledging the impact of destination diversity on both the generation and attraction of bicycling trips, Buck and Buehler (2012) estimated that an increased density of restaurants and retail establishments with a liquor license in a half-mile from a station significantly predicted station activity.

A handful of studies have further examined the urban design and transportation system determinants of bikeshare demand. Modeling trip activities between origin-destination pairings, El-Assi et al. (2017) found that an increase in the number of intersections with major roads along a modeled route decreased station-to-station activity. Measured within a station area, length of major roads had a negative relationship with both arrival and departure rates (Faghih-Imani et al., 2014) as well as destination choice of both annual members and daily customers (Faghih-Imani & Eluru, 2015). The second study (Faghih-Imani & Eluru, 2015) also noted a positive link between bikeshare use and length of minor roads, which are likely located in more residential environments and provide system users with bike-friendly facilities. Other studies have underlined this positive connection between bike infrastructure (e.g., on-street lane, off-street trail) provision and bikeshare ridership (Buck & Buehler, 2012; Wang et al., 2016).

Additionally, characteristics of the transit network have been explored in the literature because of the potential for bikeshare to offer a solution to the last-mile problem (Shaheen et al., 2010). A positive association between bikeshare station attraction (Faghih-Imani & Eluru, 2015) and station-to-station trips (El-Assi et al., 2017) with the presence of a rail station within a bikeshare station area has been established. In contrast, Bachand-Marleau et al. (2012) found that an increased bus stop density near a rider's residence decreased bikeshare system ridership, while Rixey (2013) found no statistical significance with this relationship.

The influence of the built environment surrounding a station on bikeshare ridership has received inadequate attention in the literature (Ahillen, Mateo-Babiano, & Corcoran, 2016). Previous bikeshare studies have examined the connection between the built environment and bikeshare ridership by specifying the former phenomenon as a set of independent metrics rather than an interconnected collection of various characteristics describing a physical setting. Future studies investigating the interactions between the built environment and bikeshare adoption should also develop metrics operationalized at a station area-level, as this spatial extent likely reflects the context a pedestrian or cyclist accessing a station will encounter (Rixey, 2013). Future studies of bikeshare use should also focus more on the built environment encountered by an individual reaching a bikeshare station (Rixey, 2013). This study addresses each identified research gap by exploring an extensive set of correlated built environment features within a walkable catchment area to identify a station area typology for the CaBi program serving Washington, DC metropolitan region.

3 Method

3.1 Study area and data sources

Data on the 433 CaBi station locations opened as of December 2016 and ridership activity for 2016 were collected. The largest share of stations was located in Washington, DC (55%) and Arlington County, Virginia (20%); the two jurisdictions where the 10 original stations were sited for the opening of the program on 20 September 2010 (Halsey, 2010). Since its inception, the bikeshare system has continued

its growth into the neighboring jurisdictions of Montgomery County, Maryland; Alexandria, Virginia; and Fairfax County, Virginia, which at the time of this study had 57, 31, and 19 stations, respectively. This assortment of bikeshare stations provided a substantial number of sites located in a variety of spatial contexts.

To understand this potential contextual variation and ultimately develop a system-wide bikeshare station area typology, a set of 24 built environment measures were collected. Table 1 summarizes this comprehensive list of built environment measures representing the land development pattern measures, urban design, and transportation infrastructure elements in a station area. Land development pattern measures describing the density and land-use mix of an area were provided by the 2010 US Census Bureau and 2014 Longitudinal Employer-Household Dynamic (LEHD) datasets, while 12 urban design and transportation infrastructure measures describing the station areas were collected from OpenStreet-Map (OSM).

Table 1. Built environment measures

Built Environment Measure	Description	Data Source
Population	Number of persons	Census 2010
Housing units	Number of housing units	Census 2010
Employment	Number of jobs	LEHD 2014
Retail employment	Number of retail jobs	LEHD 2014
Activity	Number of persons and jobs	Census 2010/ LEHD 2014
Employment entropy 3-class	3-class entropy index: Balance of office, retail/services/entertainment, and industrial	LEHD 2014
Employment entropy 5-class	5-class entropy index: Balance of office, retail, industrial, services, and entertainment	LEHD 2014
Employment entropy 8-class	8-class entropy index: Balance of office, retail, industrial, services, entertainment, education, healthcare, and public administration	LEHD 2014
Employment-population balance 1	Ratio of jobs to persons	Census 2010/ LEHD 2014
Employment-population balance 2	Ratio of retail jobs to persons	Census 2010/ LEHD 2014
Employment-population balance 3	Ratio of retail/service/entertainment jobs to persons	Census 2010/ LEHD 2014
Street blocks	Number of street blocks	OSM 2016
Intersections 1	Number of 3- and 4-way intersections	OSM 2016
Intersections 2	Number of 5-way and above intersections	OSM 2016
Cul-de-sacs	Number of cul-de-sac streets	OSM 2016
Street links	Number of street links	OSM 2016
Connected node ratio	Ratio of 3- and 4-way intersections to all nodes	OSM 2016
Alpha index	Ratio of observed circuits to maximum number of circuits	OSM 2016
Beta index	Ratio of street links to intersection nodes	OSM 2016
Gamma index	Ratio of observed street links to maximum number of street links	OSM 2016
Connectivity cyclomatic index	Number of route alternatives between nodes	OSM 2016
Distance to nearest rail station	Street network miles from bikeshare station to nearest WMATA metro rail station	OSM 2016
Distance to nearest bus stop	Street network miles from bikeshare station to nearest bus stop	OSM 2016
Distance to nearest park	Street network miles from bikeshare station to nearest park	OSM 2016

3.2 Built environment and station area measurement

Given the variation in which these collected built environment measures are operationalized, an apportionment and smoothing process was implemented to best ensure conformity in their spatial representation. First, an arrangement of 656-foot grid cells was casted over the DC region study area. This cell size was chosen to minimize the number of instances where where CaBi stations were in the same grid cell. Second, half-mile areal buffers were extended from the centroids of these grid cells. Those built

environment measures provided at a block geography were then intersected with these areal buffers in a geographic information system (GIS) environment. An apportionment process next allocated the measurement values to these buffers. Often, a census block would be fully contained by the areal buffer; however, in instances where this was not the case, the percent of the block's area inside the buffer served as a factor for allocating population and employment shares. This apportionment process was conducted for the land development pattern measures, as the area-based urban design and transportation infrastructure measures were calculated using disaggregate information (e.g., intersection node) or data not confined to a census block (e.g., street network).

A smoothing process was next undertaken in which the measurement totals summarized for each buffer were then assigned to the grid cell at the center of the buffer. The outcome of this two-step weighting scheme was the assignment of all secondary aggregated built environment data to a spatially uniform geography. For the three distance-based built environment measures, the shortest network distance between the cell's centroid and the nearest bus stop, rail station, and park boundary was calculated. Table 2 presents the descriptive statistics for the built environment in a subset of grid cells with a CaBi station as well as all cells in the larger study area, including: Washington, DC; Montgomery County, Maryland; Prince George's County, Maryland; Arlington County, Virginia; Fairfax County, Virginia; and the Virginia cities of Alexandria, Fairfax, and Falls Church.

Table 2. Descriptive statistics of built environment measurement at a grid cell geography

Built Environment Measure	Station Areas (n=272)		Study Area (n=10,931)	
	Mean	St. Dev.	Mean	St. Dev.
Population	1,776.38	1,504.15	358.92	582.93
Housing units	973.92	905.55	147.15	287.35
Employment	2,874.37	5,279.29	210.48	1,013.77
Retail employment	76.49	62.84	16.40	28.00
Activity	3,268.64	2,682.08	569.78	848.13
Employment entropy 3-class	0.54	0.09	0.36	0.22
Employment entropy 5-class	0.51	0.08	0.41	0.19
Employment entropy 8-class	0.59	0.07	0.45	0.20
Employment-population balance 1	133.36	268.79	15.72	76.79
Employment-population balance 2	4.08	17.25	0.73	4.74
Employment-population balance 3	51.36	115.29	7.09	34.72
City blocks	29.98	12.22	9.84	8.97
Intersections 1	62.04	36.89	13.08	16.38
Intersections 2	2.05	2.27	0.27	0.69
Cul-de-sacs	2.14	2.19	2.87	3.50
Street links	81.18	41.94	22.96	27.06
Connected node ratio	0.91	0.09	0.66	0.25
Alpha index	0.03	0.02	0.10	0.05
Beta index	2.05	0.25	1.40	0.53
Gamma index	0.70	0.08	0.49	0.17
Connectivity cyclomatic index	349.66	131.08	110.69	97.61
Distance to nearest rail station	0.83	0.78	109.33	1,024.29
Distance to nearest bus stop	0.17	0.20	0.98	1.68
Distance to nearest park	0.37	0.24	0.95	0.89

3.3 Latent class clustering analysis

Latent class clustering analysis (LCCA) is a model-based, probabilistic clustering method used to analyze the relationships among select variables, which may be observed in the original dataset or unobserved variables belonging to a latent class (Haughton, Legrand, & Woolford, 2009). As with other cluster strategies, which include heuristic procedures or k-means cluster analysis, the general impetus behind using LCCA is an identification of distinct groups of observations that are interrelated with other relatively homogenous membership categories (Fraley & Raftery, 2002; Beckman & Goulias, 2008). The LCCA model structure for continuous indicators is presented in Equation 1 (Vermunt & Magidson, 2002).

$$P(y_i|\theta) = \sum_{k=1}^K \pi_k P_k(y_i|\theta_k) \quad (1)$$

In Equation 1, y_i is the i th observation of the modeled indicator, K is the number of clusters, and π_k is the prior probability of membership to latent class cluster k . For continuous indicators θ , P_k is the density function.

When compared to more traditional clustering approaches, LCCA offers several advantages because of its model-based approach to clustering observations or, in the case of this study, bikeshare station areas. First, by using a statistical model to estimate an observation's class membership, model summary statistics such as the Bayesian Information Criterion (BIC) and Integrated Classification Likelihood (ICL)—described elsewhere (Biernacki, Celeux, & Govaert, 1998; Vermunt & Magidson, 2002)—are produced to allow for a less arbitrary standard for selecting the number of latent classes. Second, latent class models, otherwise known as finite mixture models, accommodate unscaled and unstandardized variables with mixed measurement levels, which allows LCCA model outputs to be interpreted in their original units (Higgins & Kanaroglou, 2016). Third, although each sampled object is assumed to belong to a discrete class, latent class membership accounts for uncertainty because of the calculation of posterior probabilities based on model parameters and observed scores, which permits the prediction of other objects belonging to the population that were not sampled (Vermunt & Magidson, 2002).

4 Results

Described below, existing bikeshare station areas were classified using an LCCA framework. The introduction of this bikeshare station area typology was then extended to a planning application in which the current CaBi network was expanded in a simulation exercise to identify potential new bikeshare station locations. The proposed station areas were next classified using the bikeshare typology, with the station-level ridership activity for the proposed stations subsequently being estimated.

4.1 Bikeshare station area typology

Table 3 provides the estimation results of the latent class model, which was performed in the R statistical software program (R Core Team, 2016) using the *mclust* package (Fraley & Raftery, 2007). The presented model with five latent class clusters, whose distribution within the sample is noted by the mixing probability, had the lowest BIC and ICL statistic. Observed variables in this final specification were manually chosen through a variable reduction process guided by both a diagnostic of the zero-order correlations and desire for parsimony in the built environment indicators chosen to determine class membership. The five indicators in the final model specification reflect important built environment de-

terminants of bikeshare ridership described in the literature, and pertain to land development patterns, urban design, and transportation infrastructure features. This model specification produced five latent clusters and a lower BIC statistic than the next-best six-class (BIC=-8,523; ICL=-8,587) and seven-class (BIC=-8,549; ICL=-8,590) specifications.

Table 3. Model estimation results of latent class cluster analysis

Built Environment Indicator (means)	Cluster membership				
	1	2	3	4	5
Activity	21.81	19.16	40.76	98.98	38.23
Employment-population balance 1	1.30	27.65	40.61	585.02	454.30
Beta index	2.03	1.83	2.23	2.40	1.98
Distance (miles) to nearest rail station	0.84	1.34	0.34	0.28	0.65
Distance (miles) to nearest park	0.32	0.59	0.33	0.18	0.30
<i>Cluster Summary</i>					
Number of bikeshare stations	141	60	99	71	62
Mixing probability	0.32	0.14	0.22	0.17	0.15
<i>Model Performance</i>					
Log likelihood					-4,030.63
Bayesian information criterion (BIC)					-8,449.79
Integrated classification likelihood (ICL)					-8,494.93

Estimation results indicated variation across these different aspects for one-half mile bikeshare station areas in the current CaBi system. The following class descriptions detail the built environment indicators influencing class membership in addition to providing information on the spatial distribution of these station area latent classes in the study area (Figure 1).

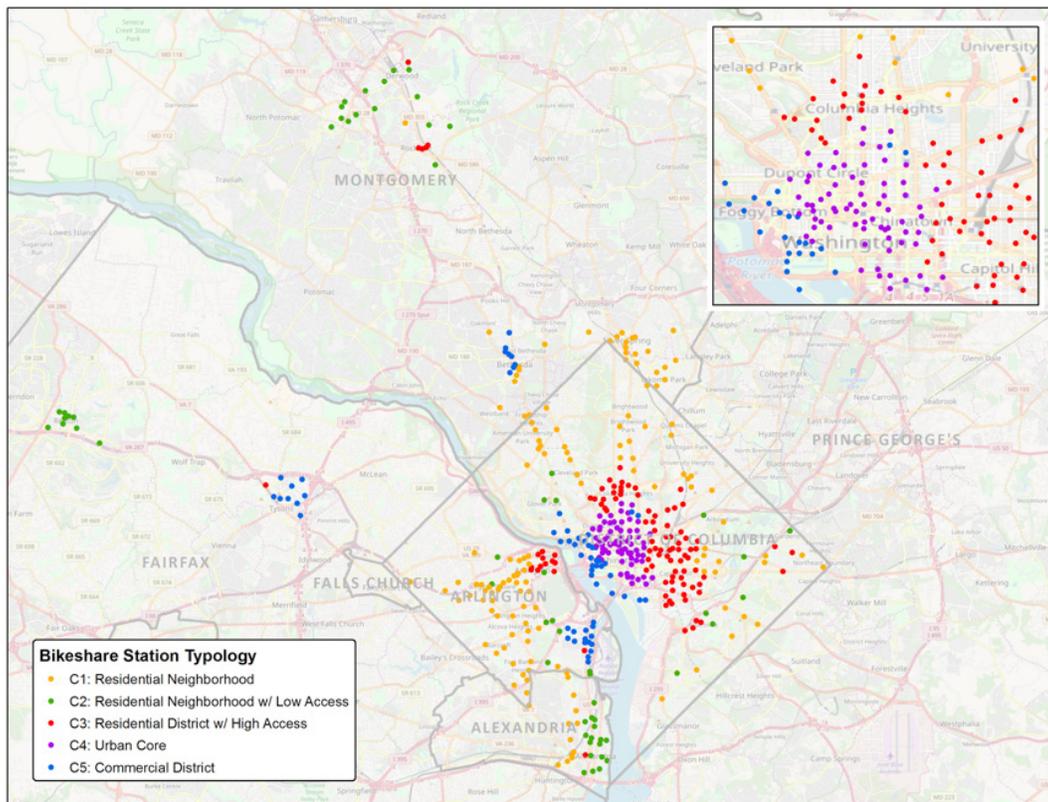


Figure 1. Bikeshare station typology estimated by latent class cluster analysis

4.1.1 Cluster 1: Residential neighborhood

One-third (33%) of bikeshare stations are located in a residential neighborhood, which is characterized by a balanced ratio of jobs-to-persons but comparatively low total in these two activity-related features. On average, CaBi stations in this built environment context are situated in a grid-based traditional street network, while being located within a half-mile from the nearest park and one mile from a rail station. Station areas classified as this latent cluster are dispersed throughout the metro region, with the exception of Fairfax County, and exemplified by the bikeshare station near the Arlington Mill Community Center at South Dinwiddie Street and Columbia Pike.

4.1.2 Cluster 2: Residential neighborhood with low access

The fewest percent of bikeshare stations (14%) are classified as this second latent class, which is described by a low activity density, non-traditional street design, and limited access to a nearby rail station or park. Bikeshare stations in this area type—while found in Alexandria—are typically situated the farthest from the District of Columbia's borders, with the CaBi station at the YMCA in Reston, Virginia serving as a model.

4.1.3 Cluster 3: Residential district with high access

There are 99 bikeshare station areas (23%) with membership in this latent cluster, which has the second highest average value for activity density and street network connectivity. Stations in this setting gener-

ally tend to be located one-third of a mile from the closest rail station and park. The built environment surrounding the CaBi station at Second Street and Massachusetts Avenue NE epitomizes this area type, which is mostly distributed within the District.

4.1.4 Cluster 4: Urban core

This cluster has the third highest share of existing bikeshare stations (16%). On average, stations within this category are located in areas with the highest activity density, most traditional network design, and greatest access to a nearby park and rail station. Unsurprising, CaBi stations found in the urban core also exhibit the highest jobs-housing ratio. Stations in this context such as the one at 16th Street and K Street NW are located immediately south of the National Mall and to its north.

4.1.5 Cluster 5: Commercial district

Only 62 bikeshare stations (14%) belong to this final latent class. Like those stations in the previous cluster, these stations are situated in activity dense contexts with disproportionately low numbers in population and close proximity to nearby parks. However, this station area tends to be characterized by a more suburban street network and greater distance to a Metrorail service. Stations within this context can also be found in Washington, DC as well as Arlington, Bethesda, and Fairfax County. The CaBi station at Greensboro and International Drive by Tyson's Galleria is an archetype station within this final category.

4.2 Planning application

The presented LCCA approach to analyzing bikeshare station areas offers several advantages to forecasting the potential location of new bikeshare stations and estimating ridership between those stations. The aggregation of key built environment measures into posterior probabilities simplifies any predictive model specification and reduces the burden of narrowing a large field of potential independent variables. Endogenous to these cluster membership probabilities is a known statistical correlation between a given set of built environment variables and bikeshare locations. To demonstrate the efficacy of using these probabilities in a predictive model, the five-class CaBi station area typology was integrated into the Bikeshare Location and Origin-destination Count (BikeLOC) model in a planning application for the Washington, DC metro region. Described in detail below, BikeLOC is a hierarchical, non-recursive model with four modules that (i) estimate built environment measures of a study area delineated into user-defined sub-units, (ii) classify the feasibility of siting a new docking station within these sub-units, (iii) perform a cluster analysis of existing and simulated bikeshare station locations to predict a final bikeshare system, and (iv) generate station-to-station bikeshare activity using an origin-destination trip count algorithm.

The Open-source Public-data Built Environment Urban Modeler (OPBEUM) module is designed to be used for any spatial boundary, defined as an exogenous input by its user, within the United States. Spanning the chosen boundary, a fishnet of grid cells is created and a variety of built environment measures are calculated in these analytic sub-units. For this application, 656-foot grid cells were selected for analyzing the current jurisdictions served by CaBi as well as Prince George's County, Maryland; Falls Church, Virginia; and Fairfax City and County, Virginia. The module used block-level LEHD employment and US Census housing and population data to calculate activity levels. Additionally, the model used a local installation of OSM to calculate each grid cell's beta index and network distance between its centroid and nearest rail station.

In the next two BikeLOC modules, a potential bikeshare station location was derived by training a Support Vector Machine (SVM) algorithm with a Radial Basis Function (RBF) kernel. The underlying

grid cell-based built environment was used as independent variables to train the LCCA algorithm with both grid cells known to contain a bikeshare station and a random sample of cells without bikeshare stations. The SVM then classifies all study area grid cells as either supportive of a bikeshare station or not. In this application, the grid cell-based built environment variables were substituted with the class membership probabilities, which trained the SVM to predict bikeshare supportive grid cells by calculating the posterior probabilities of known bikeshare locations and then using the estimated model coefficients to determine the probabilities of the remaining study area grid cells. The result was 648 out of 10,931 study area grid cells could potentially support a bikeshare station.

To reduce the number of potential station locations to only include those that form an interconnected network of stations needed to facilitate ridership, BikeLOC performs a spatial cluster analysis to eliminate those selected grid cells that are either orphaned (one station area without near neighbors), form an island (pair of station areas without nearby neighbors), or belong to sparse networks (groups of station areas that are too distant to encourage ridership). These potential locations were eliminated by first measuring the kernel density of the spatial patterns of the grid cells through an empirical distribution function. Using Fourier transformations, the distribution approximation was combined to form a discrete version of the kernel density and linearly approximated to evaluate station density across potential locations. The density distribution is then transformed to contours and the contour encapsulating a pre-defined density is selected. The contour is converted to a polygon and used to select the grid cell that will continue on in the ridership module. The results of the spatial cluster analysis on our 648 potential bikeshare station locations was a 42-percent reduction based on the density threshold, which resulted in 272 feasible bikeshare station locations (Figure 2).

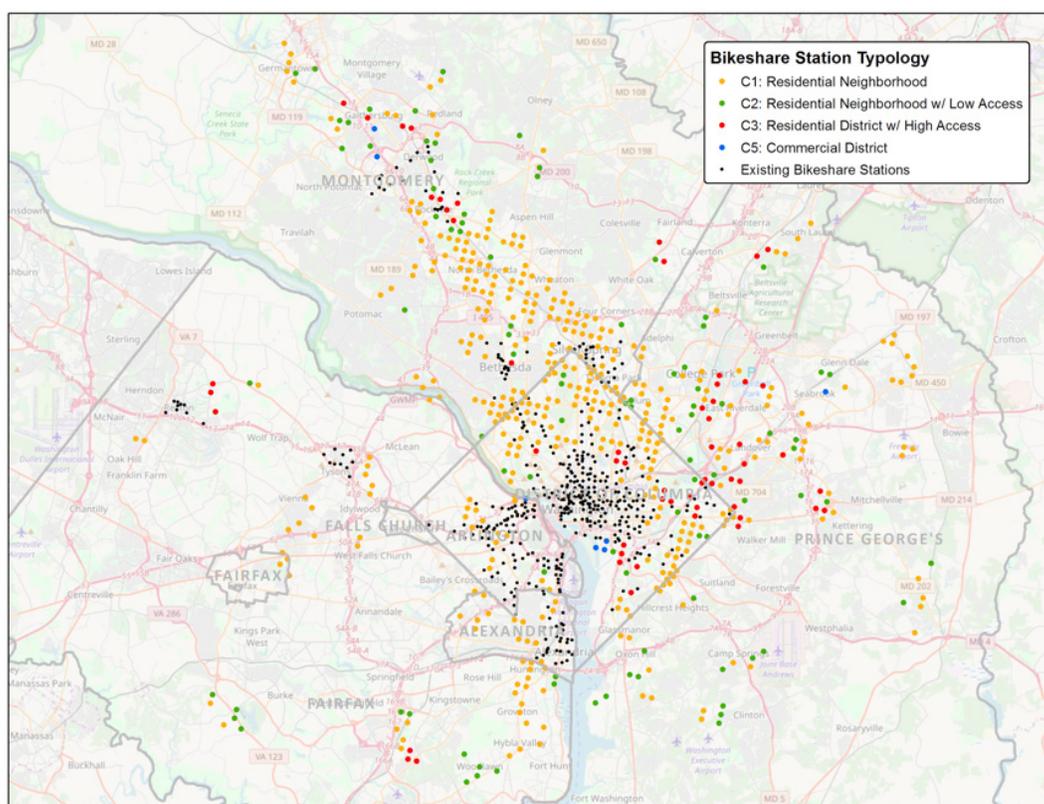


Figure 2. Simulated bikeshare station expansion produced by BikeLOC

The final BikeLOC module estimated the directional ridership between new bikeshare station locations. First, the network distance between each station pair was measured using an A* shortest path algorithm across a local optimized PostgreSQL version of the North America OSM, which established a preferred route between the two future station areas. Currently, the only specified cost function was distance; however, other bike-oriented attributes may be added in future applications. In addition to distance, the elevation change between each potential station pair was measured. The two variables are accompanied by the LCCA probabilities to complete the suite of variables specifying a second SVM designed for regression with a Gaussian RBF kernel that predicts station-level ridership between the potential pairs, which were further reduced to only include station pairs within 10 miles of one another.

In this application, the SVM was trained using 11,222,482 station-to-station CaBi trips observed from winter 2011 to spring 2016. By using a trained SVM, the annual ridership was predicted between a simulated set of 50,473 station pairings; reduced from 102,831 possible origin-destination pairings. The results of this analysis, summarized in Table 4, show an estimated 3,673,035 annual unlinked trips between 433 existing and 272 modeled stations.

Of this forecasted trip total, 62-percent of the annualized trip activity takes place between existing bikeshare stations. Data from observed station-to-station CaBi rider activity revealed the largest portion of trips occurred between stations in residential districts with high access (Cluster 3). The second highest number of forecasted trips originated and ended at stations located within a commercial district, followed by travel between stations both located in the urban core setting (Cluster 5). Of all observed trips with trip ends in different station areas, those trips originating in a residential district with high access and ending at a station within an urban core context (Cluster 4) produced the highest forecasted annual ridership. Perhaps unsurprisingly, the pairing of existing stations with the lowest ridership total was for unlinked trips starting within the urban core and ending at a station in a residential neighborhood with low access (Cluster 2); likely a product of the dispersed spatial distribution of these station areas within the metro region.

The trip distribution for newly simulated bikeshare stations produced by BikeLOC found a vast majority of trips (60.79%) to start and end at stations in a residential neighborhood (Cluster 1). With the second highest forecasted ridership activity for simulated stations occurring for trips ending at CaBi stations in this setting from stations within a residential neighborhood with low access, BikeLOC has identified an opportunity to expand CaBi service into these two station area types that are presently underserved by bikeshare but could support network expansion. As evidenced in Figure 2, the model did not simulate new stations in the urban core, primarily because this area type is relatively limited in the region and already has a strong distribution of bikeshare stations. Similarly, trip distribution totals with either trip end being located at a station in the commercial district is lower than the trip totals for observed station-to-station trips in this context because of the limited prospect for further CaBi expansion into this already well-served setting.

Table 4. Forecast unlinked annual station-level ridership of expanded bikeshare network

<i>Trip Flow: Existing Bikeshare Stations</i>		Station Area Type (Destination)				
Station Area Type (Origin)	1	2	3	4	5	
Cluster 1	227,396	19,637	111,928	45,545	34,707	
Cluster 2	19,361	38,968	14,332	7,704	31,005	
Cluster 3	97,051	9,210	345,361	154,516	105,205	
Cluster 4	25,301	3,248	121,989	235,715	113,070	
Cluster 5	18,950	21,558	90,510	110,376	260,318	
<i>Trip Flow: Simulated Bikeshare Stations</i>		Station Area Type (Destination)				
Station Area Type (Origin)	1	2	3	4	5	
Cluster 1	857,183	118,878	89,939	---	435	
Cluster 2	138,786	25,628	18,947	---	10,651	
Cluster 3	104,115	17,919	11,684	---	2,333	
Cluster 4	---	---	---	---	---	
Cluster 5	10,532	1,683	1,015	---	346	

5 Conclusion

This study has presented an innovative method for classifying bikeshare stations based on their surrounding built environment context. As dock-based bikeshare systems continue to extend their networks into previously unserved areas, further insight is needed to better understand whether certain contexts can support system expansion with high performing bikeshare stations. While prior research has sought to define what neighborhoods perform best (Ahillen et al., 2016), our study introduces a five-class bikeshare station area typology based on key built environment characteristics to identify the combination of contextual characteristics that distinguish bikeshare station locations. Additionally, as demonstrated, this proposed typology can then be integrated into a transportation planning tool to objectively inform planners and agencies of candidate locations for system expansion and provide a sense of the associated cycling activity that may result from siting in a particular physical context. Presently, data-driven approaches to siting bikeshare stations are unusual in practice.

Future research extensions should explore this application and others. Presently, the five identified clusters describe the objective environment in a half-mile areal buffer extending from the center of 656-foot grid cells with a bikeshare station; however, the sensitivity of model results to operationalizing the five indicators at different geographic scales warrants further study. Second, the addition of site-level attributes to the cluster membership model may provide greater variation across the five station areas. A National Association of City Transportation Officials (NACTO) bikeshare station siting guide (NACTO, 2016) proposed three station types based on their street-level placement: stations in the street, stations on the sidewalk, and stations in open places. The consideration of site-level measures in this study's proposed station area typology would provide richer information to planners and operators seeking to maximize a station's utility to riders. Finally, future applications should offer a deeper inspection of the modeled behavior of between-station activity and the types of riders generating and completing trips at these station area types. In all, there are a number of exciting prospects for extending the utility of these proposed station area types beyond forecasting the ridership associated with a simulated system expansion.

References

- Ahillen, M., Mateo-Babiano, D., & Corcoran, J. (2016). Dynamics of bike sharing in Washington, DC, and Brisbane, Australia: Implications for policy and planning. *International Journal of Sustainable Transportation*, 10(5), 441–454.
- Bachand-Marleau, J., Lee, B. H. Y., & El-Geneidy, A. (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, 66–71.
- Beckman, J. D., & Goulias, K. G. (2008). Immigration, residential location, car ownership, and commuting behavior: A multivariate latent class analysis from California. *Transportation*, 35, 655–671.
- Biernacki, C., Celeux, G., & Govaert, G. (1998). Assessing a mixture model for clustering with the integrated classification likelihood. (Report 3521). Rocquencourt, France: Institut national de recherche en informatique et en automatique (INRIA).
- Buck, D., & Buehler, R. (2012). Bike lanes and other determinants of Capital Bikeshare trips. Paper presented at 91st Annual Meeting of the Transportation Research Board, Washington, DC.
- Corcoran, J., & Li, T. (2014). Spatial analytical approaches in public bicycle sharing programs. *Journal of Transport Geography*, 41, 268–271.
- DeMaio, P. (2009). Bike-sharing: History, impacts, models of provision, and future. *Journal of Public Transportation*, 12(4), 41–56.
- 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(3), 589–613.
- Faghih-Imani, A., Eluru, N., El-Geneidy, A., 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. (2015). Analyzing bicycle-sharing system user destination choice preferences: Chicago's Divvy system. *Journal of Transport Geography*, 44, 53–64.
- Fishman, E. (2016) Bikeshare: A review of recent literature. *Transport Reviews*, 36(1), 92–113.
- Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458), 611–631.
- Fraley, C., & Raftery, A. E. (2007). Model-based methods of classification: Using the mclust software in chemometrics. *Journal of Statistical Software*, 18(6), 1–13.
- Halsey III, A. (2010, September 21). New bikeshare program provides wheels to casual cyclists in DC, Arlington. *Washington Post*. Retrieved from <http://www.washingtonpost.com/wp-dyn/content/article/2010/09/20/AR2010092003815.html>. Accessed March 13, 2018.
- Higgins, C. D., & Kanaroglou, P. S. (2016). A latent class method for classifying and evaluating the performance of station area transit-oriented development in the Toronto region. *Journal of Transport Geography*, 52, 61–72.
- Haughton, D., Legrand, P., & Woolford, S. (2009). Review of three latent class cluster analysis packages: Latent Gold, poLCA, and MCLUS. *The American Statistician*, 63(1), 81–91.
- Krykewycz, G., Puchalsky, C., 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, 117–124.
- Ma, T., Liu, C., & Erdogan, S. (2015). Bicycle sharing and public transit: Does Capital Bikeshare affect Metrorail ridership in Washington, DC? *Transportation Research Record: Journal of the Transportation Research Board*, 2534, 1–9.

- National Association of City Transportation Officials. (2016). *Bike share station siting guide*. Retrieved from http://nacto.org/wp-content/uploads/2016/04/NACTO-Bike-Share-Siting-Guide_FINAL.pdf
- O'Brien, O., Cheshire, J., & Batty, M. (2014). Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography*, *34*, 262–273.
- R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- 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*, 46–55.
- 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*, 159–167.
- Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J.A. Hagenaars & A. L. McCutcheon (Eds.) *Applied latent class analysis* (pp. 89–106). Cambridge: Cambridge University Press.
- Wang, X., Lindsey, G. Schoner, J., & Harrison, A. (2016). Modeling bike share station activity: Effects of nearby businesses and jobs on trips to and from stations. *Journal of Urban Planning and Development*, *142*(1), 1–9.