
https://doi.org/10.1109/TITS.2022.3145669

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it’s important to you. Thank you.
Quality of service Measure for Bike Sharing Systems

Huthaifa I. Ashqar, Mohammed Elhenawy, Hesham A. Rakha, Fellow, IEEE, and Leanna House

Abstract—Bike sharing systems (BSSs) are becoming an important part of urban mobility in many cities given that they are sustainable and environmentally friendly. BSS operators spend great efforts to ensure bike and dock availability at each station. Measuring the quality of service (QoS) of each station and/or the entire system is critical for efficient system operations. The traditionally-known QoS measure reported in the literature is based on the proportion of problematic stations, which are defined as those with no bikes or docks available to users. This measure neither exposes the spatial dependencies between stations nor does it discriminate between stations in the BSS. Hence, we propose a novel QoS measure, namely the Optimal Occupancy, in which: 1) the temporal variations in arrival and pick up rates at individual stations are considered; 2) the discriminative property of the Optimal Occupancy is demonstrated using Analysis of Variance (ANOVA) procedures; and 3) geo-statistics, which have not been used before, are applied to explore the spatial Optimal Occupancy variations and model variograms for spatial prediction. This study uses an anonymized bike trip dataset from 34 stations in downtown San Francisco to compare the traditionally-known QoS measure and the proposed Optimal Occupancy measure. Results reveal that the Optimal Occupancy is beneficial, outperforms the traditionally-known QoS measure, and produces a better prediction of the QoS at nearby locations. In addition, the Optimal Occupancy can be used to predict candidate locations for the introduction of new stations in an existing BSS.

Index Terms— Bike Sharing System, Optimal Occupancy Quality of service, Spatial Analysis, Urban Computing

I. INTRODUCTION

A growing population, with more people living in cities, has led to increased pollution, noise, congestion, and greenhouse gas emissions. One possible approach to mitigating these problems is encouraging the use of bike sharing systems (BSSs). BSSs are an integral part of urban mobility in many cities and are sustainable and environmentally friendly. As urban density increases, it is likely that more BSSs will appear due to their relatively low capital and operational costs, ease of installation, and the ability to track bikes [1].

BSS operators take great efforts to ensure bike and dock availability at each station. This task can be difficult as the movement of users is highly dynamic, difficult to predict, and redistributing bikes is expensive. Recent studies have shown that there are spatial dependencies in bike usage at different stations [2-7], and that imbalances in the spatial distribution of bikes occur due to one-way use and short rental periods [5]. Thus, it is necessary for operators to understand the spatial dependencies to more effectively manage the system. For example, operators could improve the quality of service (QoS) by identifying the best candidate spots for new stations. However, finding the best QoS measure for a station in a heterogeneous BSS and using it to study the spatial dependencies in the system is a challenging problem.

We investigated the state-of-art QoS measure and found it to be largely indiscriminative at the station level. In this study, we propose a new QoS measure, namely the Optimal Occupancy, to discriminate between different stations in heterogeneous BSSs. We demonstrate that Optimal Occupancy is not only discriminative but can also capture the spatial correlations in a BSS.

II. RELATED WORK

One of the first BSSs in the United States was established in 1964 in Portland, with 60 bicycles available for public use. Although BSSs are still relatively limited, at present many cities, including San Francisco and New York, have launched BSS programs. These programs implement different payment structures, conditions, and logistical strategies. Modeling bike sharing data is an area of significant research interest [8]. In general, the main goals of previous studies have been to boost the redistribution operation [9-13], to gain new insights into and correlations between bike demand and other factors [14-19], and to support policy makers and managers in making optimized decisions [5, 14, 20, 21].

Research questions that have been studied previously include the strategic design, operation, and analysis of BSSs [8]. Due to the potential benefits to operators, measuring the quality of service of stations or the entire system [22] has become an
appealing research topic. In some cases, operators measure the fraction of time that their stations are full or empty as a measure of the QoS of the system [23]. Fricker et al. considered the limiting probability that a station is empty or full as the performance measure. They argued that the optimal proportion of bikes at a station is slightly more than half the capacity of a station in a homogeneous system. In a heterogeneous system, however, they concluded that this performance metric collapses due to the heterogeneity nature of the system [24].

Lin and Yang [25] studied bike station measures of service. They argued that the measures of service quality in the system should include two measures: the availability rate, which was defined as the proportion of pick-up requests at a bike station that are met by the bicycle stock on hand, and the coverage level, which is the fraction of the total demand at both origins and destinations that is within some specified time or distance from the nearest rental station. Fricker and Gast [26] proposed a stochastic model of a homogeneous BSS and investigated the impact of users’ random choices on the number of problematic stations. Problematic stations were defined as stations that, at a given time, have no bikes available or no available docks for bikes to be returned to. Consequently, the performance of the system was determined by the proportion of problematic stations. However, these measures have critical drawbacks: (1) as BSSs usually offer two services: picking up bikes, and returning bikes; some measures fail to take into account the QoS of returning bikes to stations; (2) some of the studies assume that, in contrast to real systems, the system is homogeneous; and (3) while some studies modeled the system as heterogeneous, they failed to consider the variability of the system parameters (i.e., arrival and pickup rates) throughout the same day or across the different days of the week and their dependency on the individual station.

In any BSS, one of the keys to success is the location and distribution of bike stations [8, 25]. Some studies have worked on locating bike stations using different methods, such as location-allocation models [27-30], and an optimization method that maximizes the demand covered and takes the available budget as a constraint [31]. The spatial distribution of the potential demand is a fundamental element in optimal location modeling [32]. In order to estimate the potential demand, several studies used preference surveys to evaluate both the factors influencing the use of the bicycle mode and choice of routing [33-36]. Potential demand has also been estimated by considering the population, employment associated with each building, and the number of trips generated for each transport zone [27, 32, 37]. However, there are some limitations and drawbacks in the methods previously used to find the optimal station location: these methods are basically used to plan new systems and might be difficult to predict new stations in existing systems; they are aimed at serving the local population on selected days (e.g., workdays) and certain places in the studied area (e.g., large parks) have neither population nor jobs and yet may attract a considerable number of trips. In this study, we used spatial analysis (i.e., geo-statistics) as described in [38, 39] to explore the spatial configuration of an existing BSS to predict the location of new stations. Although spatial analysis was used in some transportation applications such as urban planning [40, 41], crash analysis [42, 43], and travel modeling [44], this study is the first attempt to use spatial analysis in predicting the locations of new stations in BSSs.

This paper makes two major contributions to the literature: (1) we propose a new discriminative QoS measure that reflects the spatial dependencies in a heterogeneous BSS capturing the variability in arrival and pickup rates; and (2) we use this QoS measure with geo-statistics to model a spatial variogram that could predict the QoS in nearby areas for the purpose of locating new stations in an existing BSS.

### III. Proposed QoS Measure

BSSs are highly heterogeneous; i.e. the arrival rates, pickup rates, origins, and destinations between stations in diverse areas and topographies are very different. These parameters may also vary with the time of day, day of the week, and season [45]. In this study, we consider that the bike sharing system has \( N \) stations, in which each station \( i \) may have a unique capacity \( C_i \) (i.e., maximum number of docks). We assume that the dynamics of the system are as follows. Users reach the stations to pick up a bike at varying departure rates \( D_i \) at station \( i \) (we named it departure rate as their intention is to take the bike and depart to their destination stations). This departure rate \( D_i \) varies throughout the day for different days of the week. If there are no available bikes, the user leaves the system or waits until another user arrives to return a bike. Users arrive at their destination stations to return a bike at a varying rate \( A_j \) at station \( j \). Similar to the departure rate, the arrival rate \( A_j \) depends on station \( j \) and varies throughout the day for different days of the week. If there are less than \( C_j \) (i.e., capacity) bikes in the station, the user returns the bike and leaves the system. If the station is full, the user either chooses another station to return the bike or waits until another user reaches the station to pick up a bike.

To consider the impact of system heterogeneity, we introduce a new QoS measure for each station, namely the **Optimal Occupancy**, in which the variations in arrival and pick up rates are captures. The Optimal Occupancy of a station is formulated in terms of two services: (1) picking up bikes, and (2) returning bikes. As each station \( i \) has a finite number of docks (i.e., capacity), two thresholds should be defined. The lower threshold \( (L_i) \) is the point when the number of available bikes \( B_{i,t} \) in station \( i \) at time \( t \) drops low enough that the probability of a user not finding a bike is very high. The upper threshold \( (U_i) \) is the point when the number of bikes \( B_{i,t} \) in a station \( i \) at time \( t \) is high enough that the probability of a user not finding a dock to return a bike is very high. For example, if a station’s capacity is 25 docks and the number of available bikes at time \( t \) is within [5, 20], then the station is considered functional; otherwise it needs to be rebalanced (i.e., it is a problematic station). In that sense, the Optimal Occupancy \( (O_{op}) \) is formulated as the ratio of the total time that a station is functional \( (f_t) \) during a given interval to the duration of the interval \( (t_{total}) \) as:
\[ O_{oi} = \frac{t_{i,f}}{t_{i,total}} \]  
where \( t_{i,f} = \sum_{t=0}^{t_{i}} X_i(t) \)  
where \( X_i(t) \) is the status function and defined as:  
\[ X_i(t) = \begin{cases} 1, & \text{station } i \text{ is functional} \\ 0, & \text{station } i \text{ is problematic} \end{cases} \]
and station \( i \) is functional if \( B_{i,t} \in [L_i, U_i] \) at any given time \( t \).  

As the two thresholds \( L_i \) and \( U_i \) define the functionality of the station, \( L_i \) and \( U_i \) are correlated with the departure rate \( D_i \) and arrival rate \( A_i \), respectively. Both \( D_i \) and \( A_i \) randomly vary throughout the day, for different days of the week, and different months of the year. In this study, we assume that \( D_i \) and \( A_i \) vary only over different days of the week (DoW), and different months of the year (M) to be consistent with the length of the study interval (see Analysis and Results section). In fact, \( D_i \) and \( A_i \) are the bike counts picked up (\( D_i \)) or returned (\( A_i \)), respectively, per unit time (\( t_{i,total} \)). In that sense, and to reflect the stochastic phenomenon in the system, \( D_i \) and \( A_i \) were modeled using a Poisson Regression Model (PRM) with an exposure variable. In a previous study [46], we found that PRM is an appropriate method to model the bike counts at a specified time \( t \) because the counts are discrete, non-negative integers, and represent the number of available bikes at a specified time at each station in the network. In PRM, each observation \( i \) is allowed to have a different mean \( \mu_i \), where \( \mu_i \) is estimated from empirical data. The PRM assumes that \( D_i \) and \( A_i \) have a Poisson distribution, and its logarithm (i.e., link function) can be modeled as a linear combination of parameters.

Following are brief descriptions of these two models. Exposure is a measure of how the bike counts are divided. Since both rates are bike counts per unit time, time is considered as the exposure. The model contains a \( \log(t_{i,total}) \), which is called the offset variable, as a term that could be added to the regression coefficients:

\[ D_i \text{ or } A_i \sim \text{Poisson}(\theta_{i}^{D(\text{or } A)}) \]  
where \( \theta_{i}^{D(\text{or } A)} = \frac{\mu_i}{t_{i,total}} \) and \( \theta_{i}^{A} = \frac{\lambda_i}{t_{i,total}} \)  
\[ \log\left(\frac{\mu_i}{t_{i,total}}\right) = \beta_0 + \beta_1 \text{DoW} + \beta_2 M \]  
\[ L_i = \mu_i \]  
\[ U_i = C_i - \lambda_i \]  
where \( \mu \) and \( \lambda \) are the mean of the Poisson distribution for picked up bikes and returned bikes, respectively at each station \( i \).

In that sense, problematic stations can be redefined as stations that, at any given time \( t \), have fewer bikes available than the expected bike counts to be picked up during the analysis discretization period or more bikes than the difference between the station capacity and the expected bike counts to be returned during the analysis discretization period. The next sections in this study will further explain the concept of the proposed Optimal Occupancy QoS measure by applying it to a real BSS dataset and comparing the new definition of problematic stations with the one previously used.

### IV. Dataset

In 2013, San Francisco launched the Bay Area Bike Share System (now called the “Bay Wheels” BSS), a membership-based system providing 24-hours-per-day, 7-days-per-week self-service access to short-term rental bicycles. Members can check out a bicycle from a network of automated stations, ride to the station nearest their destination, and leave the bicycle safely locked for someone else to use [47]. The Bay Area BSS is designed for short, quick trips, and as a result, additional fees apply to trips longer than 30 minutes. In this system, bike stations connect users to transit, businesses, and other destinations in four areas: downtown San Francisco, Palo Alto, Mountain View, and downtown San Jose [47]. The Bay Area BSS is available to everyone 18 years and older with a credit or debit card. The system is designed to be used by commuters and tourists alike, whether they are trying to get across town at rush hour, traveling to and from the Bay Area Rapid Transit (BART) and Caltrain stations, or pursuing daily activities [47].

![Fig. 1. Stations map. [47]](image-url)
V. ANALYSIS AND RESULTS

In a BSS, the QoS measure should reflect the spatial dependencies of BSS stations in addition to describing the performance of a station’s service. Consequently, we investigated the traditionally-known QoS measure using the Bay Area BSS dataset in San Francisco. We found that it was neither satisfying in exposing the spatial dependencies between stations nor adequate in describing the performance of the service.

The first QoS measure presented in different studies (such as in [23, 24, 26]) is that of problematic stations, defined as stations that, at a given time, have no bikes available or no available spots for bikes to be returned to. This definition has been mainly used to describe the overall performance of the system. However, we used that definition to find a QoS measure for a specific station by computing the ratio of the total time that a station is not problematic during a given interval to the length of the interval. The second measure is our proposed QoS measure, Optimal Occupancy ($O_{op}$), which redefines problematic stations as stations that, at any given time $t$, have fewer bikes available than the expected bike counts to be picked up during the analysis discretization interval or more bikes than the difference between the station capacity and the expected bike counts to be returned during the analysis discretization interval. Similarly, we used our definition to find the Optimal Occupancy for a specific station by computing the ratio of the total time that a station is not problematic (i.e., functional) during a given interval to the length of the interval. For this specific dataset, and to effectively represent the service in the system, we defined the length of the study interval in both definitions as running from 8 a.m. to 5 p.m., which was found to be the peak hours for the system [50]. Fig. 2 shows the locations of the stations with the corresponding results of the two QoS average measures (over two years) of 34 stations in the Bay Area Bike Share in San Francisco. The measures were first found for every 15 minutes at each station then averaged over the interval of the peak hours for the system.

A. Analysis of Variance (ANOVA)

Analysis of variance (ANOVA) was used to determine whether there are any statistically significant differences between the means of the two QoS measures. Before interpreting the results of the hypothesis tests, we checked the ANOVA assumptions, and the hypothesis test results were found to be trusted. BSSs are highly heterogeneous, with arrival rates and pickup rates between stations in diverse areas and topographies varying with the time of day, day of the week, and season [45, 50]. Therefore, to fairly compare the two measures, we compared the daily values for specific months and days. ANOVA was used to analyze the differences among four group means for all 34 stations: (1) Tuesdays of February, (2) Tuesdays of July, (3) Mondays of February, and (4) Mondays of July. The $p$-values resulting from testing the groups of traditionally-known QoS measures are 0.7704, 0.8400, 0.5099, and 0.7443, respectively. This means that the Null Hypothesis is true and there are no significant differences ($p > 0.05$). On the other hand, the $p$-values resulting from testing the groups of the proposed QoS measures ($O_{op}$) are 2.73E-29, 3.25E-36, 7.34E-41, and 1.42E-30, respectively. This means that the Null Hypothesis is rejected and that there are significant differences between the measures of the stations ($p < 0.05$). Fig. 3 shows the differences among the Tuesdays of February group means for all 34 stations. It clearly demonstrates that the traditionally-known QoS cannot be used to discriminate between the stations, while the proposed Optimal Occupancy is discriminative to a sufficient extent. In that sense, recognition of the differences between the QoS of stations is not required in and of itself but because it is necessary for operators to effectively manage the system and it appears to reflect the dynamics of the BSS. Although we present the results of only four groups, in fact we examined the ANOVA test for other groups that cover most of the days of the week and months of the year. The results were found to be consistent with the results presented here.
B. Spatial Analysis

We applied geo-statistics to explore the spatial configuration of Optimal Occupancy variations. We used two packages in R, geoR to analyze geostatistical data [51] and gstat to perform geostatistical modelling and prediction [52]. The analysis was performed to assess whether the proposed Optimal Occupancy measures can reflect the spatial dependencies and be used to predict the QoS in nearby areas. This will allow operators to determine candidate spots for new stations in the BSS, which will increase the overall QoS of the system.

Fig. 2. ANOVA test for Tuesdays of February for the 34 stations for (a) traditionally-known QoS, and (b) proposed QoS.

Fig. 3. The locations, and the values of the (a) proposed QoS, and (b) traditionally-known QoS measures.
Spatial statistics attempt to develop inferential methods to properly account for the spatial dependences in the presence of georeferenced observations. Spatial modeling typically contains a specification of a mean function and a model of the correlation structure (i.e., variogram), which is a description of the spatial continuity of the data. The variogram is the key function in geostatistics as it is used to fit a model of the spatial correlation of the observed phenomenon [38, 53]. A variogram model is chosen by plotting the empirical variogram, which is a simple nonparametric estimate of the variogram, and then comparing it to various theoretical shapes available. A variogram could be mathematically defined as [38].

\[ \gamma(\Delta x, \Delta y) = \frac{1}{2} \varepsilon [(Z(x + \Delta x, y + \Delta y) - Z(x, y))^2] \] (9)

where \( Z(x, y) \) is the value of the variable of interest at location \( (x, y) \), and \( \varepsilon [\cdot] \) is the statistical expectation operator.

The variogram, \( \gamma(\cdot) \), is a function of the separation between points \( (\Delta x, \Delta y) \), and not a function of the specific location \( (x, y) \). However, one common assumption of the spatial analysis is that it is isotropic. An isotropic variogram means that the correlation between any two observations depends only on the distance between those locations and not on their relative direction; otherwise, it is anisotropic [39].

A series of directional empirical variograms (including directions between 0° and 180°) was investigated to highlight the main observations’ directions and check the spatial isotropy in the proposed QoS data. The results illustrate that we cannot assume isotropy and that the directional empirical variogram for 45° outperforms other variograms as it reflects the correlation between the observations and the distance. The empirical variogram for 45° using transformed coordinates was estimated and is illustrated in Fig. 4. It shows a steady increase in the semi-variance over increasing distance intervals to an absolute maximum between 1.0 and 1.5 km. For greater distances, Fig. 4 displays an oscillatory state with a second maximum around 2.5 and 3 km.

![Fig. 4. The empirical variogram for 45° using transformed coordinates.](image)

Modeling variograms are usually used for spatial prediction (i.e., interpolation). Most practical studies used Exponential, Spherical, and Gaussian models. As we assumed anisotropy, we applied the maximum likelihood estimation of spatial regression models to estimate the angle for geometric anisotropy of the three models. The Exponential variogram model yields the most beneficial realization of the spatial process in the BSS. While the Spherical model yields a decent estimation, the Gaussian model fails to fit a variogram that manifests the spatial correlation. We also applied the maximum likelihood estimation for the same three models to fit the traditionally-known QoS measure to compare it with the Optimal Occupancy. Similarly, the Exponential variogram model outperforms the Spherical and the Gaussian models. Results in TABLE 1 show some inferences. We used Bayesian Information Criterion (BIC) to evaluate the performance of the two models. BIC is an information theoretic approach, which can be used to assess model error via the estimated loss of information associated with a candidate model [54]. BIC score is defined as a function of the likelihood of the fitted model and the number of parameters in the model, in which the second part serves to penalize the model fit according to its complexity. BIC addresses the model evaluation problem where more parameters lead to increased likelihood. To resolve this, BIC penalize additional parameters [54]. According to the BIC of the spatial and non-spatial models, the spatial model for Optimal Occupancy outperforms the non-spatial one, but the traditionally-known QoS non-spatial model outperforms the spatial one. This shows that the traditionally-known QoS cannot expose the spatial dependencies between stations. Therefore, using Optimal Occupancy is more advantageous than using the traditionally-known QoS. As the BIC for the spatial model demonstrates, Optimal Occupancy as a measure is more gainful and would result in better prediction of the QoS in a BSS.

<table>
<thead>
<tr>
<th>Parameters Estimation of the Exponential Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Optimal Occupancy</td>
</tr>
<tr>
<td>Traditionally-known QoS</td>
</tr>
</tbody>
</table>

C. Optimal Location for New Stations

We proposed the Optimal Occupancy as a QoS measure to: (1) allow the operator to keep track of the performance of different stations in a BSS, so for example they may increase the number of docks/available bikes in a station; (2) identify the optimal location of new stations in existing systems using a data-driven decision management approach. In the previous section, geo-statistics was used to model a spatial variogram that could predict the QoS in nearby areas for the purpose of locating new stations in an existing BSS. The model was used to produce new QoS datasets in order to build a QoS surface for the case study area. Fig. 5 shows the QoS surface for the case study area in San Francisco. This surface could be used to quantify and visualize the QoS measures represented by contours in the surface. Looking at the surface in Fig. 5, there are four hot spots (red-colored) that could be considered as candidates to add new stations nearby or increase the number of docks in a station. By doing so, we convert the surface into a more homogeneous QoS terrain, demonstrating a more functional BSS (i.e. less problematic stations at any given time) and easier to rebalance. It is also interesting to mention that during our study, Ford GoBike, the operator of the case study BSS, has added different coming-soon stations near the
abovementioned spots or added more docks to others. For example, a coming-soon station is to be built very close to Station 50 shown in Fig. 2(a), which we hypothesize was added to increase the functionality of Station 50 (see [55]).

In [46, 56, 57], a model was developed to predict the bike counts at each station in the Ford GoBike system using Random Forest (RF) as a univariate regression algorithm for different prediction horizons. Modeling bike counts using RF produced a Mean Absolute Error (MAE) of 0.37 bikes/station, which means the model was found to be promising. Station 50, namely Harry Bridges Plaza Station, was also found to be one of the stations that are highly unpredictable due to the high fluctuations in bike counts. When the area around Harry Bridges Plaza Station was studied, it was hypothesized that this high incoming/outgoing demand comes from it being an open air area at the end of a market with numerous restaurants nearby, where artists, skaters, tourists and others congregate to enjoy the happenings and beautiful scenery [58]. In that sense, we will use the developed model in [56, 57] to prove our hypothesis that if a new station is added, for example near Station 50, it will increase its functionality. We will compare the proposed QoS values for two different days of the week, Monday and Tuesday of July, before and after adding the new suggested station near Station 50. The model was used to predict the bike counts at Station 50 every 15 minutes for each of the selected days to estimate the proposed QoS using Equations (1) through (8). We assumed that the new station will cover only a third of the two types of services that Station 50 used to serve. The resulting QoS for Station 50 was improved after adding the new suggested station by increasing from Optimal Occupancy from 0.52 to 0.84 and from 0.43 to 0.79 for Monday and Tuesday of July, respectively.

It is worthwhile mentioning that a low Optimal Occupancy in a station may have two reasons, namely, too few or too many bikes in the station. Although this was not the case in the abovementioned example, increasing the capacity of the station due to a low Optimal Occupancy can only mitigate the issue of having too few bikes in the station as the system is assumed to be heterogeneous. However, we argue that this could be determined by investigating the ratio of the available bikes to the capacity of the station during the period of analysis and consequently identifying whether to increase the capacity or decrease it.

![Fig. 5. Predicted QoS surface for the case study area.](image)

VI. CONCLUSION

This study proposes a new QoS measure for stations in a BSS that are operating at Optimal Occupancy. The Optimal Occupancy at a station is formulated in terms of two types of services: (1) picking up of bikes and (2) returning of bikes. It is formulated as the ratio of the total time a station is functional during a given time interval to the length of the interval. Consequently, we redefined problematic stations as stations that, at any given time, have fewer bikes available than the expected bike counts to be picked up during the analysis discretization period or have more bikes than the difference between capacity and the expected bike counts to be returned during the analysis discretization period.

We further studied the concept of the proposed QoS measure by applying it to a real dataset of 34 stations in the San...
Francisco Area and also compared the new definition of problematic stations with the one previously used. First, results from ANOVA analysis clearly demonstrate that the traditionally-known QoS measures cannot be used to discriminate between BSS stations, whereas the Optimal Occupancy is found to be sufficiently discriminative. Recognition of the differences between the QoS of stations benefits the effective management of the system and appears to reflect the dynamic nature of the BSS.

In addition, we applied geo-statistics to explore the spatial configuration of the Optimal Occupancy variations and model variograms for spatial prediction. The empirical variogram shows a steady increase in semi-variance over increasing distance intervals to an absolute maximum between 1.0 and 1.5 km. The Exponential variogram model was fit and yielded the most beneficial realization of the spatial process in the BSS. Results revealed that the spatial model for Optimal Occupancy outperforms the non-spatial one. Furthermore, Optimal Occupancy as a measure is more informative and would result in better prediction for the QoS in nearby locations. However, the spatial model was used to produce new QoS datasets in order to build a QoS surface for the case study area. Adding new stations nearby the hot spots in the surface, we could convert the surface into a more homogeneous QoS terrain, demonstrating a more functional and easier to rebalance BSS as a result of this change. For example, the resulting QoS for Station 50 was improved after adding the new suggested station from 0.52 to 0.84 and from 0.43 to 0.79 for Monday and Tuesday of July, respectively.

There are some ideas that could serve as future steps to this study. First, this study did not use data from different geographical locations in the U.S and the number of stations was relatively small. A different dataset from different locations in the U.S. that has a larger number of stations could further improve the veracity of this study. Second, this research relied on a dataset of a dock-based bike sharing system, which means it might not be directly applicable to dockless BSSs. However, the dock-based system is still being used in different cities including the system that we used the data from in the San Francisco Bay Area, and the Capital Bikeshare system that serves the Washington DC Metropolitan area with about 500 stations. A different dataset from a dockless bike sharing system could further improve the veracity of this study. Third, although we recognize the critical role of potential demand, this was not possible given limitations in the data. The results of this study will primarily provide important new insights to policymakers, operators, and researchers, who spend a great amount of time and effort to satisfy users and allow operators to better manage their BSSs.

VII. REFERENCES


Mohammed Elhenawy received the Ph.D. degree from Virginia Polytechnique Institute and State University under the Transportation Research Laboratory. He is currently a Senior research fellow at the Centre for Accident Research and Road Safety—Queensland (CARRS-Q) and a Faculty Member at Queensland University of Technology. He is also working on Ipswich Connected Vehicle Pilot Project and also worked on glare on tunnel endpoints: road safety problem, new methodological approach for analyses and simulations, and hold the red evaluation. His research interests include data science, artificial intelligence, and intelligent transportation systems.

Hesham A. Rakha (M’04, SM’18, F’20) received the Ph.D degree from Queen's University, Kingston, Ontario, in 1993. He is currently the Samuel Reynolds Pritchard Professor of Engineering in the Department of Civil and Environmental Engineering and the Department of Electrical and Computer Engineering (Courtesy) at Virginia Tech, and the Director of the Center for Sustainable Mobility at the Virginia Tech Transportation Institute. His research focuses on large-scale transportation system optimization, modeling, and assessment. He works on optimizing transportation system operations, including vehicle routing, developing various network and traffic signal control algorithms, developing freeway control strategies (speed harmonization and ramp metering), and optimizing vehicle motion (lateral and longitudinal control of connected automated vehicles (CAVs)) to enhance their efficiency and reduce their energy consumption while ensuring their safety. Dr. Rakha is a Fellow of the IEEE. He is the recipient of IEEE ITS Outstanding Research Award (2021) from the IEEE Intelligent Transportation Systems Society. He was the author/co-author of six conference best paper awards, namely: 19th ITS World Congress (2012), 20th ITS World Congress (2013), VEHITS (2016), VEHITS (2018), and TRB (2020); received the most cited paper award from the International Journal of Transportation Science and Technology (IJTST) in 2018; and received 1st place in the IEEE ITSC 2020 UAS4T Competition. In addition, Dr. Rakha received Virginia Tech's Dean's Award for Outstanding New Professor (2002), the College of Engineering Faculty Fellow Award (2004-2006), and the Dean's Award for Excellence in Research (2007). He is an Editor for Sensors (the Intelligent Sensors Section), an Academic Editor for the Journal of Advanced Transportation, a Senior Editor for the IEEE Transactions of Intelligent Transportation Systems, and an Associate Editor for the Journal of Intelligent Transportation Systems: Technology, Planning and Operations and the SAE International Journal of Sustainable Transportation, Energy, Environment, & Policy. Furthermore, he is on the Editorial Board of the Transportation Letters: The International Journal of Transportation Research and the International Journal of Transportation Science and Technology.

Leanna House received the Ph.D. degree in statistics from Duke University, Durham, NC, USA, in 2006. She is an Associate Professor of Statistics with Virginia Tech (VT), Blacksburg, VA, USA, where she has been since 2008. Prior to VT, she was with the Battelle Memorial Institute, Columbus, OH, USA, and subsequently served as a Post-Doctoral Research Associate for two years with the Department of Mathematical Sciences, Durham University, Durham, U.K. She has authored or co-authored over 25 journal papers and has been a strong statistical contributor to successful grant proposals, including NRT-DESE: UrbComp: Data Science for Modeling, Understanding, and Advancing Urban Populations, Usable Multiple Scale Big Data Analytics Through Interactive Visualization, Critical Thinking with Data Visualization, Examining the Taxonomic, Genetic, and Functional Diversity of Amphibian Skin Microbiota, and Bayesian Analysis and Visual Analytics.