Transportation Research Part A 96 (2017) 14–28

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

# Models for estimating zone-level bike kilometers traveled using bike network, land use, and road facility variables

Ahmed Osama, Tarek Sayed *, Alexander Y. Bigazzi

*Department of Civil Engineering, The University of British Columbia, 6250 Applied Science Lane, Vancouver, BC V6T 1Z4, Canada*

ARTICLE INFO

*Article history:*
Received 10 June 2016
Received in revised form 2 October 2016
Accepted 21 November 2016
Available online 21 December 2016

ABSTRACT

Despite the increase in studies that link bike trips with various correlates, research gaps remain, including a lack of empirical tools to predict bike kilometers traveled (BKT) using comprehensive measures. The present study evaluates the impacts of network indicators, land use, and road facility on BKT by developing zone-level ridership models. Land use and road facility data were collected for 134 traffic analysis zones (TAZs) in the City of Vancouver, Canada. In addition, graph theory was used to obtain bike network indicators for each TAZ, including measures of connectivity, continuity, linearity, slope, and length of the bike network. A full Bayesian approach, accounting for spatial random effects among the TAZs, was used to develop the models. The results suggested that more connected, dense, flat, continuous, recreational, and off-street bike networks yielded higher BKT. Models that accounted for spatial effects were found to have better fit than those that did not incorporate spatial effects, which implies the importance of considering spatial effects while modeling BKT at the aggregate level. The models provide insights about the factors that influence BKT and information about the spatial variability of the bike travel within a city, which can be useful for planning bike networks.



## 1. Introduction

Many road authorities are promoting urban cycling to support sustainability initiatives, foster public health, and reduce auto demand on congested streets. Expanding bike facilities, traffic calming, and accommodations for cyclists at intersections are among the transportation network-related policies usually adopted to encourage cycling (Buehler and Dill, 2016). Numerous studies have identified associations between cycling rates and bike infrastructure development (Nelson and Allen, 1997; Dill and Carr, 2003; Parkin et al., 2008; Buehler and Pucher, 2012). In addition to the infrastructure effects, various bike ridership models have shown that bike trips are positively associated with other variables such as proximity to universities, flat terrain, employment density, and land use mix (Haynes and Andrzejewski, 2010; Ryan et al., 2010; Griswold et al., 2011; Strauss and Miranda-Moreno, 2013).

Despite various studies linking cycling levels and bike counts at individual locations to a wide range of covariates, there are no models of bike travel volume that can account for the combined impact of bike and roadway network characteristics and land use at a zonal level. Modeling bike kilometers traveled (BKT) at a zonal level can provide information that isolated bike counts and mode share data cannot capture. Count, mode share, and BKT data can have different spatial patterns and different relationships with network configuration variables. Variation in BKT by zone reflects the aggregate spatial

* Corresponding author.
*E-mail addresses:* ahmed.osama@eng.asu.edu.eg (A. Osama), tsayed@civil.ubc.ca (T. Sayed), abigazzi@civil.ubc.ca (A.Y. Bigazzi).

http://dx.doi.org/10.1016/j.tra.2016.11.016

distribution of bicycle trip ends (origins and destinations) and route choices. Higher BKT is expected in zones with more bicycle trip ends, longer bicycle trips, and infrastructure that attracts bicyclists on inter-zonal trips. BKT, like count data, does not distinguish between induced and diverted travel, but BKT has the advantage of accurately reflecting net changes in travel volume, whereas count data suffer from incomplete coverage.

Zonal BKT modeling can provide new intra-urban cross-sectional statistical evidence of the factors that most influence cycling volume within a city. In addition, zonal BKT models estimate the spatial variability in cycling levels, which is useful for strategic planning of active transportation investments and promotional programs. Spatial variability in cycling levels can be more broadly related to accessibility, income, and health inequities. Zonal BKT estimates can be analyzed along with travel quantities by other modes (i.e., vehicle kilometers traveled and transit passenger kilometers traveled) to investigate mode substitution not only by the number of trips but also by the cumulative travel distance. BKT models can be used in the estimation of exposure to crash risk, physical activity levels, and air pollution inhalation. In addition, zonal BKT can be used to visualize cycling levels and public engagement during strategic planning efforts.

The present study aims to develop zone-level bike ridership models that quantify the impacts of bike network, land use, and road facility variables on BKT. City of Vancouver's bike network is characterized as links and nodes, and graph theory measures are used to obtain network indicators for the 134 traffic analysis zones (TAZs) within the city. Land use and road facility data are also collected for the different TAZs. Full Bayesian analysis is undertaken accounting for spatial random effects among the TAZs. The significant associations between the different variables and BKT are then investigated to propose suitable recommendations to encourage cycling. Such recommendations can be used by transportation and planning organizations to target investments for optimum impact on existing and potential cyclists.

## 2. Literature review

### 2.1. Correlates associated with cycling levels

Several cross-regional studies found a positive association between the size and quality of the bike network and bike commute share. Nelson and Allen (1997) found that each additional mile of bikeway per 100,000 residents was associated with about 0.07% increase in bike commuters across 18 US cities. Dill and Carr (2003) used a dataset of 50 cities and found that each additional mile of facility, per square mile of city area, was associated with 1% increase in bike commuting. Buehler and Pucher (2012) also found out that a 10% increase in supply of bike lanes was associated with 3.1% increase in the number of bike commuters per 10,000 residents. Klobucar and Fricker (2007) developed a bike network analysis tool to assess the network-wide level of service offered to cyclists. The tool used route length and measures of perceived safety to quantify the bike friendliness of a street network. Ryan et al. (2010) showed that cycling as a mode of transportation was correlated with regional cycling patterns and the quality of the bike network.

Building a network that provides direct connections with minimal detours was also perceived as important by cyclists because they were shown to be highly sensitive to distance (Handy and Xing, 2011). Discontinuities in the bike network may have negative consequences such as forcing a cyclist into mixed traffic or a lengthy detour (Schoner and Levinson, 2014). Rietveld and Daniel (2004) modeled the bike share use in 103 cities in the Netherlands using several measures of infrastructure quality to conclude that the number of stops and hindrances per kilometer on any given trip was negatively associated with the bike mode share, which means that cyclists are sensitive to the speed and directness of their routes. The utility of a dedicated infrastructure is also closely related to the level of connectivity it provides (Schoner and Levinson, 2014). Schoner and Levinson (2014) collected bike infrastructure maps for 74 US cities to evaluate the impact of network structures on bike commuting rates. They used linear regression models after controlling for the demographic variables and the size of the city. The models showed the importance of directness and connectivity in predicting bike commuting rates. Mekuria et al. (2012) analyzed "low stress" networks and demonstrated the importance of connectivity through evaluating the quality of routes holistically by their weakest link rather than using a connectivity index, which is a quantitative indicator for network connectivity. Winters et al. (2016) developed a bike score that represented bikeability in 24 US and Canadian cities. They used a bike score that included bike lane score, hill score, and destination and connectivity score. They found that the correlation between mean bike score and mean journey to work cycling mode share was moderate ($r = 0.52$) at the city level, whereas at the census tract level, the correlation was 0.35.

Furthermore, various studies investigated the impact of different road facilities on the cycling levels. Wilkinson (1994) recommended a network of bike lanes, separated paths, bike boulevards, and local streets to calm auto traffic and encourage cycling. Parkin et al. (2008) found a positive significant association between the proportion of the off-street bike routes and ridership. Furthermore, Caulfield et al. (2012) found that off-street trails and green lanes increased the chances of a route being chosen by cyclists relative to a traditional bike lane. Marshall and Garrick (2011) showed that higher road classes (i.e., arterials and collectors) were less friendly and perceived as less safe by cyclists. Other studies showed that the presence of on-street parking reduced the utility of the bike lanes (Tilahun et al., 2007; Sanders, 2014). Fagnant and Kockelman (2015) used Seattle metropolitan area cyclist count data from 251 locations to develop a direct demand model for estimating peak-period cyclist counts and cycling-relevant roadway conditions. The model results showed the greatest practical significance for intersections, curb-lane width (both are associated with higher counts), and roadway speed (associated with lower counts).

Lastly, different studies have discussed the relationship between traffic, socio-demographics, and land use variables and cycling levels. Ryan et al. (2010) studied bike and pedestrian activity in San Diego County and its relationship to land use, transportation, safety, and facility type. They showed that cycling as a mode of transportation was correlated with adjacent land uses and regional cycling patterns. Schneider and Stefanich (2015) used census tract data in Wisconsin, USA, to show that the bike commute share was positively associated with several socioeconomic and local environment characteristics, including more households without automobiles, more people born in other states and countries, higher population density, being located closer to a university, more housings constructed before 1940, and higher bike facility density.

### 2.2. Graph theory

Graph theory provides techniques for evaluating network quality and measuring its impact on travel behavior. Garrison and Marble (1962) were the first to introduce graph theory principles to transportation geography. Kansky (1963) presented indices that characterized network connectivity and complexity. Graph theory measures have been applied to the field of transportation planning in several studies (Xie and Levinson, 2007; Derrible and Kennedy, 2009; Rodrigue et al., 2009; Quintero et al., 2013; Quintero-Cano et al., 2014; Osama and Sayed, 2016). However, not too many studies used graph theory measures to explain the travel behavior of nonmotorized road user. Dill and Voros (2007) found significant differences between connected node ratios and people who biked during a certain period. Network quality and connectivity were also evaluated at a micro level in the previous studies by investigating the individual discontinuities in on-street bike facilities (Krizek and Roland, 2005; Birk and Geller, 2006; Barnes and Krizek, 2005). Berrigan et al. (2010) explained individual non-motorized behavior by measuring the link-node ratio and other graph indices for a local street grid within short buffers around survey respondents' home addresses.

Although many studies have identified infrastructure, land-use, and socio-demographic correlates of bike counts or mode shares, there has been little analysis of bike travel volumes (i.e., BKT) and intra-urban variability. To the best of our knowledge, this is the first paper to evaluate the association of network, land use, and facility variables with cycling volumes at a zonal level.

## 3. Data collection

### 3.1. Data sources

In the present study, zone-level models were developed to predict BKT for 134 TAZs in the City of Vancouver. The models included explanatory variables that represented the bike network, road facility, and land use. Data from three main sources were compiled using ArcGIS:

1. Translink, the Metro Vancouver transportation authority, provided the 2013 geocoded files of the bike network, road network, and TAZ boundaries. Bike network included all types of bike facility (bike lanes, paths, etc.), whereas road network included all the road facilities (i.e., local, collector, arterial, freeway roads) in the City of Vancouver.
2. City of Vancouver provided the land use zonings and contour map (1-m resolution) of the city.
3. Acuere Analytics provided the Vancouver Cycling Data Model (VCDM). The VCDM used bike counts between 2005 and 2011 to estimate the annual average daily bike (AADB) traffic in 2011 over the City of Vancouver bike network (El Esawey et al., 2015), as illustrated in Fig. 1. The available data covered more than 810,000 hourly volumes over 7 years. The model was efficient in estimating the AADB traffic on most links of the bike network (3180 links or more than 70% of the network).

### 3.2. Analysis variables

Table 1 shows the set of variables included in the analysis. The dependent variable is BKT, calculated using the VCDM. The VCDM provides bike volumes on most of the bike network links, which are then multiplied by the corresponding link length to obtain link BKT. Link BKT is then aggregated (summed) for each TAZ. This process is not without limitations. The VCDM was efficient at estimating cycling volumes for 70% of the bike network in the City of Vancouver; the estimated BKT is therefore subject to some uncertainty. However, this method is deemed more accurate than other speculated methods for calculating zonal BKT (e.g., multiplying bike mode shares in a TAZ by the average bike trip lengths) because it incorporates real count data from links throughout the network. BKT as the dependent variable in our models is subject to measurement error, which can increase the standard errors of estimated models, but will not bias the results assuming BKT is uncorrelated with the independent variables. The independent variables in Table 1 are divided into three main categories: network indicators, land use, and road facility.

Bike network indicators include the following six variables:

- Length (L) is calculated by summing the length of all the bike network links within each TAZ after using identity function in ArcGIS to split the links among the different zones.

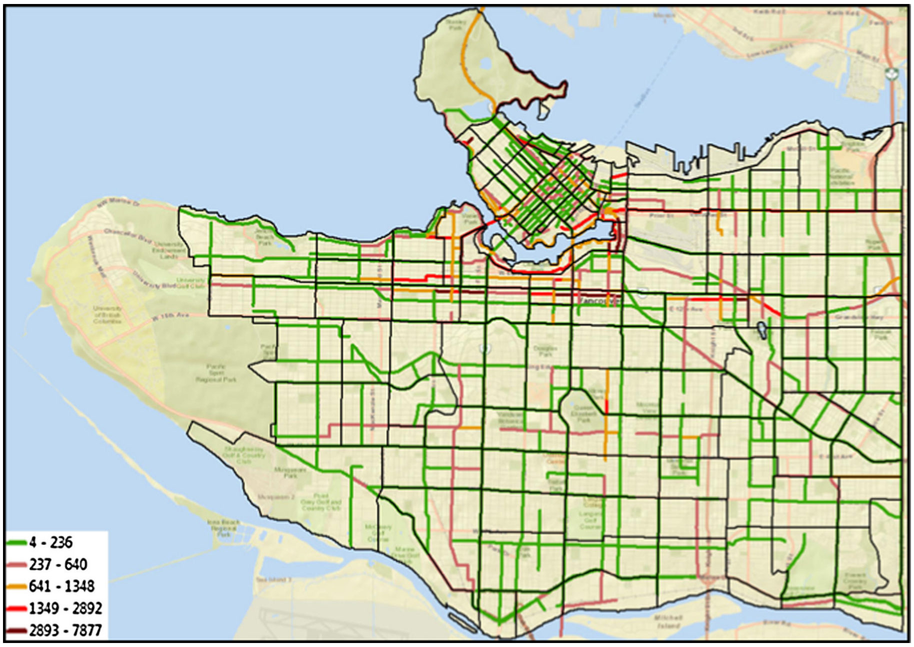



**Fig. 1.** VCDM on the TAZs.

**Table 1**
Data summary statistics.

| Variable | Description | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| BKT | Thousand BKT | 1.047 | 2.109 | 0 | 21.46 |
| *Network Indicators* | | | | | |
| L | Total Length of Bike Network Links in TAZ (km) | 3.37 | 2.52 | 0 | 17.40 |
| Conn | Degree of Bike Network Connectivity | 0.38 | 0.11 | 0 | 1 |
| Cov | Degree of Bike Network Coverage | 0.34 | 0.19 | 0 | 1 |
| AvgEdLen | Average Edge Length (Continuity) | 0.13 | 0.05 | 0 | 0.57 |
| Lin | Linearity of the Bike Network | 0.97 | 0.08 | 0.84 | 1 |
| WSlope | Average Weighted Slope for Bike Network | 2.52 | 0.90 | 0.63 | 6.65 |
| *Land Use* | | | | | |
| Res | Sum of Residential Zoned Areas in TAZ ($km^2$) | 0.35 | 0.33 | 0 | 1.59 |
| Comm | Sum of Commercial Zoned Areas in TAZ ($km^2$) | 0.033 | 0.035 | 0 | 0.21 |
| Rec | Sum of Recreational Zoned Areas in TAZ ($km^2$) | 0.13 | 0.39 | 0 | 3.66 |
| Ar | Total Area of the TAZ ($km^2$) | 0.87 | 0.74 | 0.052 | 5.31 |
| *Road Facility* | | | | | |
| Art | Total Length of Arterial Roads in TAZ (m) | 1678.12 | 1270.72 | 0 | 5745.53 |
| Coll | Total Length of Collector Roads in TAZ (m) | 1397.59 | 1256.76 | 0 | 7185.40 |
| Loc | Total Length of Local Roads in TAZ (m) | 8250.05 | 6530.40 | 0 | 30777.46 |
| Fwy | Total Length of Freeways in TAZ (m) | 1523.34 | 795.50 | 0 | 2463.97 |
| On_St | Proportion of bike network links that are on-street | 0.88 | 0.17 | 0.06 | 1 |

- Degree of Connectivity of the bike network (Conn) represents the ratio between the actual number of bike links in a TAZ and the maximum possible number of bike links within the same TAZ (Kansky, 1963). According to graph theory, the maximum possible number of bike links within a planar graph is calculated using Eq. (1), where $n$ is the number of bike nodes in the graph (Garrison and Marble, 1962). It should be noted that bike network links are joined using nodes that represent any type of intersection such as traffic lights, pedestrian crossings, give ways, and so on.

$$l_{\max} = 3(n - 2) \tag{1}$$

The value of Conn is bounded between 0 and 1. A completely connected network will have a Conn equal to 1, whereas a completely disconnected network will have a Conn equal to 0. The Conn indicator has been used in previous studies for evaluating transit networks (Derrible and Kennedy, 2011); (Quintero et al., 2013). However, two deficiencies are noticed upon applying that indicator to the study of bike networks. First, the equation that is used to calculate $l_{\max}$ should not be the same for all network shapes; rather, to estimate an accurate value for $l_{max}$, the equation should vary from one graph to another according to the network shape and nodes' configuration. Moreover, Eq. (1) is not valid when the number of nodes is less than 3. Therefore, a new indicator that can better represent the bike network connectivity is suggested.

The indicator is called the Degree of Network Coverage (Cov). Cov, shown in Eq. (2), evaluates the bike network coverage of road network (Yigitcanlar and Faith, 2010). It assumes that $l_{max}$, shown in Eq. (1), is the total number of road links within a zone. This assumption is more practical than the one in Eq. (1) because the maximum possible number of bike links shall not exceed the number of the road links within a zone. However, this indicator can be calculated only if the complete layer of the road network is available, which is the case in the present study, and if not, the Conn variable could be used. Fig. 2 shows the City of Vancouver bike network along with heat maps for both Conn and Cov by TAZ.

$$Cov = \frac{Number\ of\ Bike\ Links\ in\ TAZ}{Number\ of\ Street\ Links\ in\ TAZ} \tag{2}$$

- The continuity indicator represents bike facility length without interruptions or hindrances. For calculating bike routes continuity, a previous study by Scheltema (2012) proposed a manual methodology by counting every crossing along the key bike routes. However, this methodology is inconvenient for macro-level studies. Therefore, continuity was represented in the present study as the Average Edge Length (Kansky, 1963) shown in Eq. (3), which is the ratio of the total length of the bike network to the number of the bike links in each TAZ.

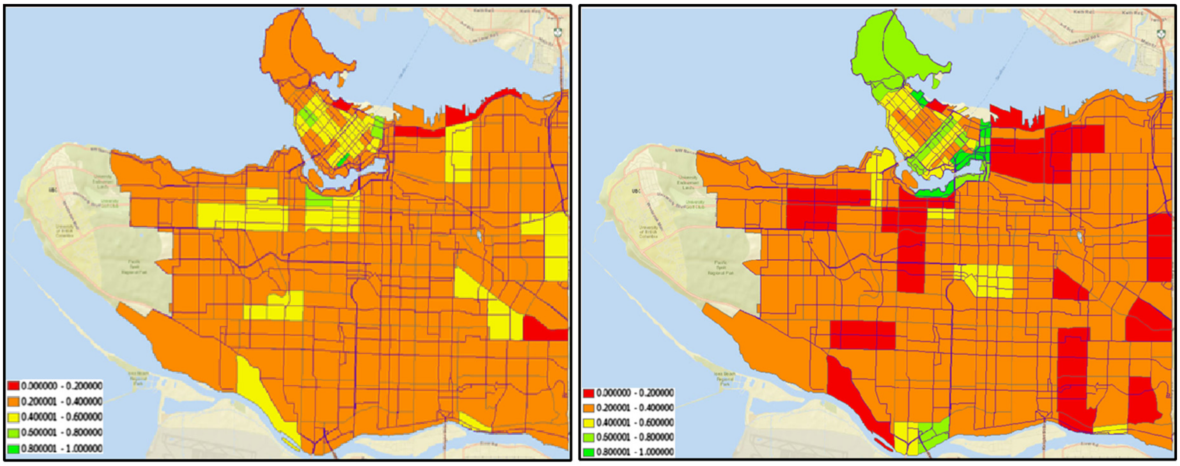
$$AvgEdLen = \frac{Total\ Length\ of\ Bike\ Network\ in\ TAZ}{Number\ of\ Bike\ Links\ in\ TAZ} \tag{3}$$

- Linearity within a bike network was also previously quantified by Scheltema (2012) on a micro-level scale as the ratio between the effective straight line and the crow line. In this macro-level study, linearity was calculated using Eq. (4) as the ratio between the modified bike network length and the original bike network length in the TAZ, where the modified bike network is a hypothetical network in which all links are drawn straight (maintaining the original nodes). A lower value of Lin represents more nonlinearity in the existing network. Fig. 3 shows the difference between the straight and nonstraight links, where nonstraight links represent any curved or irregular link.

$$Lin = \frac{Total\ Length\ of\ the\ Modified\ Bike\ network\ in\ TAZ}{Total\ Length\ of\ the\ Bike\ Network\ in\ TAZ} \tag{4}$$

To modify the nonstraight links to straight ones, the bike network was exported from ArcGIS to AutoCAD software to manage the separation of the nonstraight links from the straight links and modify the nonstraight links into straight ones. Then, the network was imported back to ArcGIS to aggregate the length of the modified network links and assess the linearity of each TAZ using Eq. (4). Fig. 4 shows the nonstraight links within the City of Vancouver bike network after being imported from AutoCAD.

- The Average Weighted Slope of the bike network in each TAZ was calculated according to the following steps. First, the absolute slope percentages along each bike link were averaged to compute the average slope of each link using the contour map of the City of Vancouver, as shown in Fig. 5. Next, the slope at each link was given a weight relative to its length.



(a) (b)

**Fig. 2.** (a) Connectivity of the bike network for City of Vancouver TAZs. (b) Coverage of the bike network for City of Vancouver TAZs.

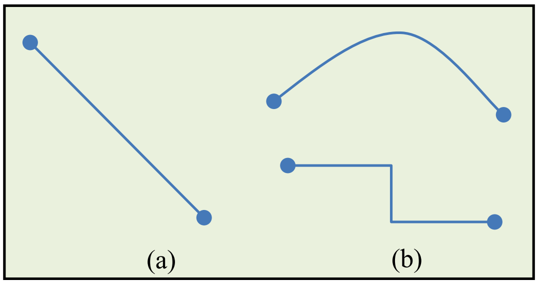



**Fig. 3.** Linearity (a) straight links (b) nonstraight links (curved or irregular).

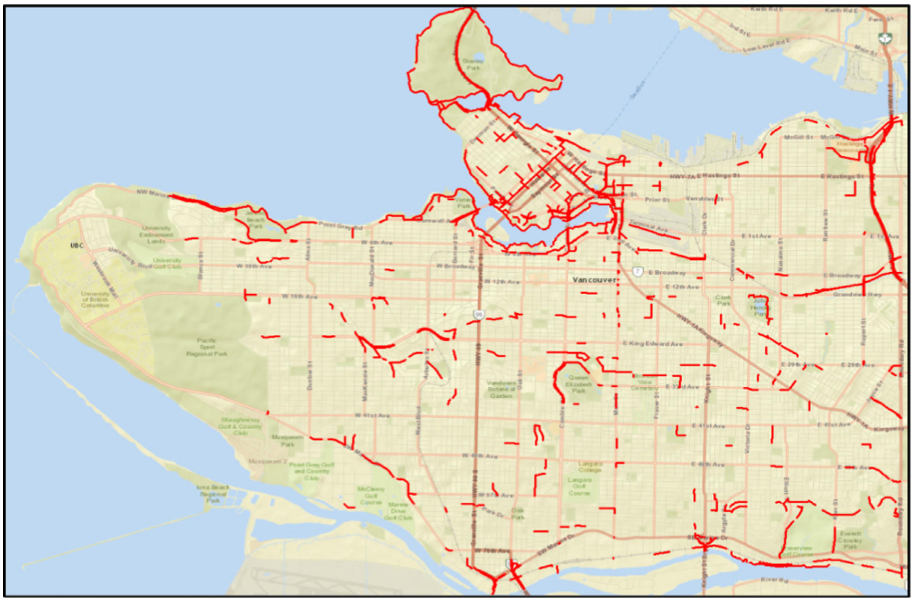


**Fig. 4.** Nonstraight links in the City of Vancouver bike network.

Finally, the average weighted slope of the bike links was calculated for each TAZ, as shown in Eq. (5), where $l$ represents the link length, $s$ represents the link's slope, and $n$ represents the number of links in the TAZ.

$$Average\ Weighted\ Slope\ in\ TAZ = \frac{\sum_1^n l_1 * s_1 + l_2 * s_2 + \cdots + l_n * s_n}{\sum_1^n l_1 + l_2 + \cdots + l_n} \tag{5}$$

For the land use category, the areas of commercial, residential, and recreational zonings were aggregated for every TAZ to obtain the total area of each land use type. Ideally, actual land use data could be obtained for future work as zonings do not always represent the actual land use. Fig. 6 shows the residential and recreational zoned areas within the City of Vancouver.

For the road network category, the total length of each road class (freeway, arterial, collector, and local roads) was aggregated (summed) for the different TAZs. Fig. 7 shows the link-based road class inventory for the City of Vancouver. Arterial-Collector proportion (ArtColl) was calculated by dividing the total length of the arterial and collector roads by the total length of the entire road network, as shown in Eq. (6).

$$ArtColl = \frac{Total\ Length\ of\ (Arterials + Collectors)}{Total\ Length\ of\ Road\ Network} \tag{6}$$

The length of the on-street bike links was aggregated for each TAZ, and then the On-Street and Off-Street bike link proportions were calculated as shown in Eqs. (7) and (8), respectively. On_St represents the proportion of the bike facilities, which are shared with the road, out of the whole bike network; whereas Off_St represents the proportion of the separated bike facilities.

$$On\_St = \frac{Total\ Length\ of\ On-Street\ Bike\ Links\ in\ TAZ}{Total\ Length\ of\ Bike\ Links\ in\ TAZ} \tag{7}$$

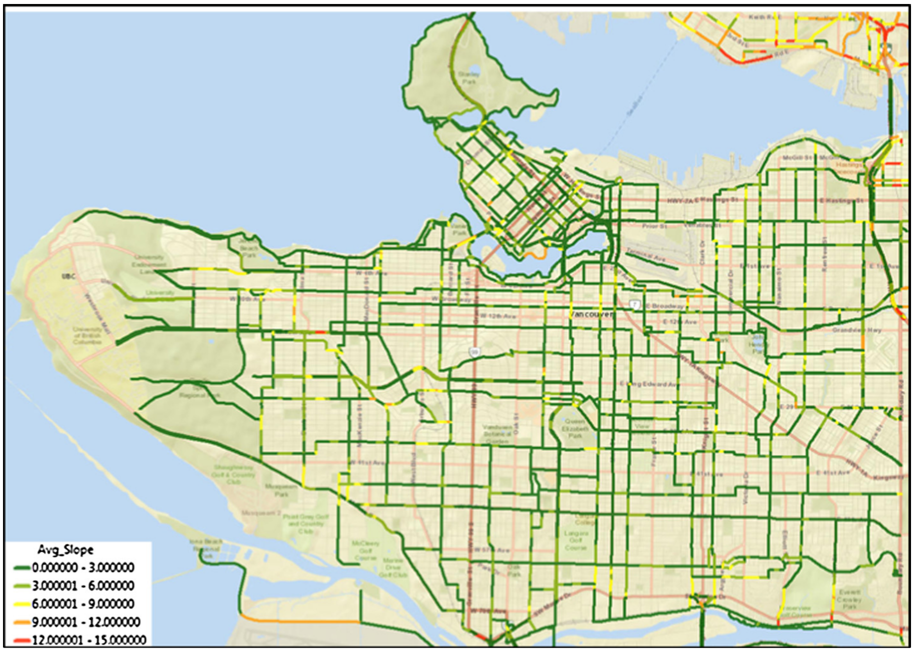
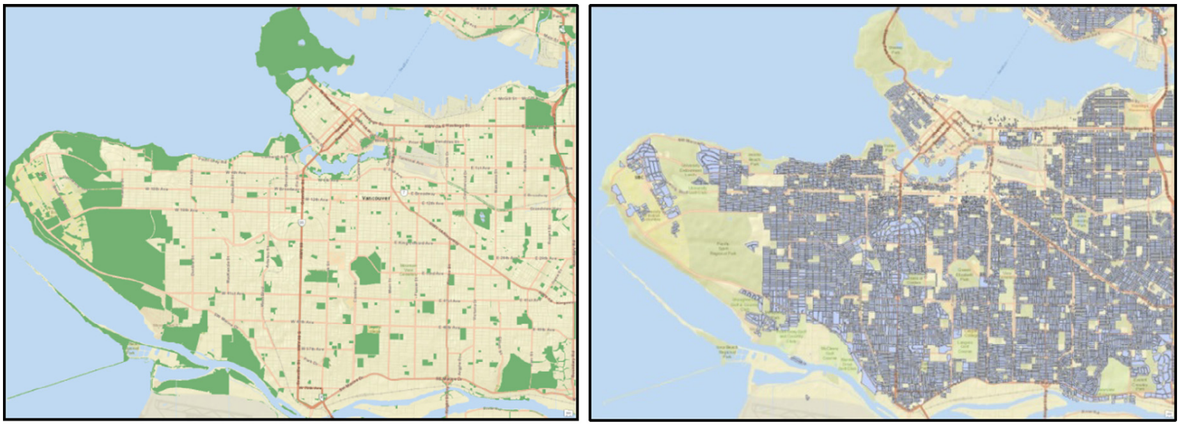



**Fig. 5.** Average slopes of the bike network links.



(a) (b)

**Fig. 6.** (a) Recreational and (b) residential areas in the City of Vancouver.

$$Off\_St = 1 - (on\_St) \tag{8}$$

Lastly, it should be noted that many variables, other than those mentioned above, were investigated during the study. However, many of those variables were found insignificant in the BKT models because of the correlation between them and other variables in the models, the insufficient sample size, or the weak relationship with BKT. Accordingly, they were not included in the analysis. Those variables included employment density, commercial area density, household density, and degree of connectivity.

## 4. Methodology

### 4.1. Linear regression models

The models in this subsection acted as a reference point for developing the full Bayes (FB) models afterward. In the present study, modeling using ordinary least squares normal regression was found to better fit the data than other types of

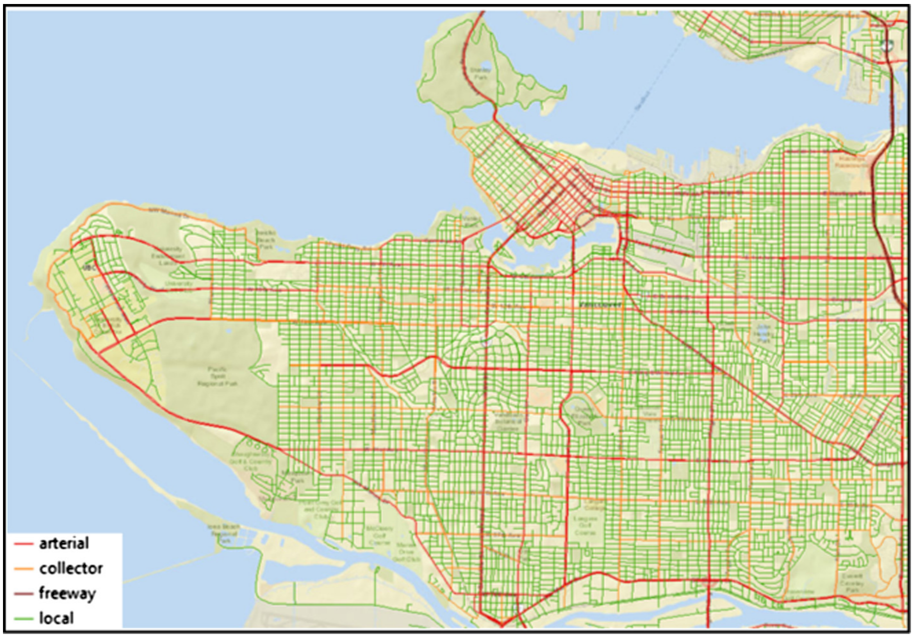



**Fig. 7.** City of Vancouver road classes.

regression (i.e., lognormal regression). Initially, three categories of macro-level linear regression models were developed separately, including the network, land use, and facility variables. The significance of the variables was evaluated using p-values, and the models' goodness-of-fit was evaluated using adjusted R-Squared.

After separately testing the impacts of network, land use, and facility variables on BKT, combined models were developed to combine the variables of the three categories to yield better predictability of BKT. The combined models included at least one variable from each studied category. The procedure of selecting the independent variables for the combined model is a forward stepwise procedure. Whether to add a variable to the model or not is decided based on two conditions. First, the added variable should be significant (at a level higher than 90% at least). Second, the variable should improve the model goodness of fit (R-Squared or deviance information criterion (DIC)).

### *4.2. Full Bayes models*

Recently, Bayesian analysis using FB hierarchical statistical models has become more popular because of its flexibility and its ability to use prior information, which results in improving the accuracy of the parameter estimates. Moreover, FB analysis can provide more accurate measures of uncertainty on the posterior distributions of the parameters' estimates (El-Basyouny and Sayed, 2009). This ability is an advantage of the FB approach over the frequentist approach because the latter does not consider uncertainty in the correlation structures, which sometimes yields an overestimation of the precision of the parameter estimates associated with the covariates. Bayesian analysis was also found to be more suitable for spatial models because of its ability to implement complex correlation structures (Aguero-Valverde and Jovanis, 2008).

#### *4.2.1. Model specification*

FB normal regression models were developed at a macro level for the present study. FB models could handle over-dispersion in the data and account for random effects. First, BKT in zone *i* was assumed to follow a normal distribution with error ε, which is considered a random error accounting for the within-zone variability. $a_0$ is the intercept value, $X_i$ are the explanatory variables, and $b_i$ are the model parameters. $u_i$ is a random error term to account for heterogeneity among TAZs. This parameter is related to the site-specific attributes, accounting for between-zone variability, and follows a normal distribution as shown in Eq. (9).

$$BKT_i = a_0 + \sum_i b_i X_i + u_i | u_i \sim normal(0, \sigma_u^2) \tag{9}$$

Spatial models can then be formulated as shown in Eq. (10) by adding $s_i$, which is a spatially correlated random effect for zone *i*, suggesting that the random error for the sites that are close to each other should be correlated.

$$BKT_i = a_0 + \sum_i b_i X_i + u_i + s_i \tag{10}$$

The spatial effects in the present study were accounted for by Gaussian CAR techniques (El-Basyouny and Sayed, 2009). Let $n_i$, $C(i)$, $\sigma_s^2$, and $S_{-i}$ represent the number of neighbors of zone $i$, the set of neighbors of zone $i$, the spatial variance, and the set of all spatial effects except $S_i$, respectively. Therefore, $S_i$ can be formulated according to Eq. (11).

$$S_i|S_{-i} \sim normal\left(\bar{s}_i, \frac{\sigma_s^2}{n_i}\right),\ \bar{s}_i = \sum_{j \in C(i)} \frac{S_j}{n_i} \tag{11}$$

Eq. (11) is based on an adjacency-based proximity measure. The conditional variance is inversely proportional to the number of neighboring zones, and the conditional mean is the mean of the adjacent spatial effects.

#### 4.2.2. Parameter estimation

Obtaining FB estimates requires the specification of prior distributions for the parameters, which reflect the prior knowledge about the parameters under consideration. The prior may be informative or uninformative depending on the availability of the prior information. The diffused normal distribution is the most commonly used prior to estimate the regression parameters (El-Basyouny and Sayed, 2009). This distribution has a zero mean and large variance. In the present study, a diffused normal distribution was used as the prior to estimate the regression parameters. For the dispersion parameter, $\sigma_u^2$, the commonly used prior is a gamma distribution with precision parameters ($\tau$, $\tau$), and 0.001 was used as the value of $\tau$. For the Gaussian CAR models developed in the present study, the prior distribution of $\sigma_s^2$ was assumed a gamma distribution with parameters ($1 + \Sigma l_i/2$, $1 + n/2$), where $l_i$ is the term contributed by each zone and is calculated by Eq. (12).

$$l_i = n_i s_i (s_i - \bar{s}_i) \tag{12}$$

The Markov chain Monte Carlo (MCMC) technique available in the WinBUGS tool was used to sample the posterior distribution and estimate the parameters. MCMC methods could sample from the joint posterior distribution repeatedly. This technique generates sequences (chains) of random points, the distribution of which converges to the target posterior distributions. A subsample is used for monitoring convergence and then excluded as a burn-in sample. Parameter estimation, performance evaluation, and inference were obtained by the subsequent iterations.

Two chains were used to run each model in WinBugs, and 20,000 MCMC iterations were discarded as burn in samples (El-Basyouny and Sayed, 2009). The chain convergences were thoroughly checked to ensure that the posterior distribution was found, which indicated that the parameter sampling should begin. Convergence can be checked in several ways. First, two or more parallel chains with diverse starting values are tracked so that full coverage of the sample space is ensured. The Bro oks–Gelman–Rubin statistic is also used to check the convergence of multiple chains, where convergence occurs if the value of the Brooks–Gelman–Rubin statistic is less than 1.2 (El-Basyouny and Sayed, 2009). Moreover, convergence can be checked by the visual inspection of MCMC trace plots of the model parameters. Finally, the ratios of the Monte Carlo errors relative to the respective standard deviations of the estimates can be calculated as a measure of convergence. As a rule of thumb, convergence occurs when these ratio values are less than 0.05.

After attaining convergence, another 20,000 iterations were performed for each chain. The summary statistics of the developed model were then estimated, and the significance of the parameter estimates was tested at the 95% level using credible intervals (El-Basyouny and Sayed, 2009).

#### 4.2.3. Comparison of models with and without spatial effect

The purpose of performing the spatial analysis was to model the spatial correlations across zones. Spatial dependence can be a surrogate for unknown and relevant covariates, thereby improving model estimation. The effects of spatial correlation were calculated by computing the spatial variation proportion of the total random effects variation according to Eq. (13).

$$\Psi = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_u^2} \tag{13}$$

where $\sigma_s^2$ is the marginal variance of spatially correlated random effects, which can be directly estimated from the posterior distribution of $s$, and $\sigma_u^2$ is the variance due to the uncorrelated random effects among the zones. Significant spatial correlation exists when $\Psi$ is found to be greater than 0.5 (Aguero-Valverde and Jovanis, 2008). It should be noted that if the variance of the random error $\varepsilon$ is added to the denominator of Eq. (13), the $\Psi$ value will decrease to 0.45 which would still show that the spatial effects account for a considerable share of the total random error. Models with and without spatial effects can be compared on the basis of the DIC. As a goodness-of-fit measure, DIC is a Bayesian generalization of Akaike's information criterion, which penalizes larger parameter models. According to Spiegelhalter et al. (2002), differences of more than 10 might rule out the model with the higher DIC. Differences between 5 and 10 are substantial. If the difference in DIC is less than 5 and the models make different inferences, then it could be misleading just to report the model with the lower DIC.

Previous studies have found that including spatial correlation would affect the parameter estimates by making some variables nonsignificant, although those variables were significant in the models without spatial effects (Karim et al., 2013). Moreover, the regression coefficient of a highly significant variable might change because of the incorporation of spatial effects. To check these two issues, the parameter estimates of the models with and without spatial effects were checked.

## 5. Results

The next three subsections present the results for the BKT models using network, land use, and facility variables separately. The fourth subsection includes models that combine all the aforementioned variable categories. Models that separate the impact of each category on BKT are presented first before developing the combined models to evaluate the prediction power of each category and its association with BKT solely. Moreover, it is interesting to study the impact of spatial effects on the significance of the variables in each category separately. Results from three different modeling techniques are discussed in each subsection: linear regression, FB normal distribution, and FB normal distribution accounting for spatial effects.

### *5.1. Network indicators models*

Two models were included in this category as it was difficult to incorporate all the network indicators in one model because of the correlations among some of the included variables (e.g., Average Edge Length is moderately correlated with Linearity; and Weighted Slope is weakly correlated with Coverage). The off-street bike lanes can also be a property of the bike network, but it was not added to the bike network models because of its correlation with some of the bike network indicators such as Average Edge Length and Weighted Slope. For example, off-street bike links usually have better continuity, so Off-Street Proportion would be highly correlated with Avergae Edge Length, resulting in multicollinearity within the model.

Model 1 included three significant variables: Coverage, Average Edge Length, and Length. The linear regression model yielded an adjusted R-squared of 0.30 and showed positive association between all the variables and BKT, as shown in Table 2. Model 2 also included three significant variables: Weighted Slope, Linearity, and Length. Linear regression analysis for this model yielded an adjusted R-squared of 0.28, as shown in Table 3. Weighted Slope is found to be negatively associated with BKT, contrary to both Linearity and Length.

The results for Average Edge Length and Coverage are plausible and in line with a previous study by Schoner and Levinson (2014), in which they showed that connectivity and directness were positively associated with bike commuting rates. In addition, the results for the Weight Slope and Length are intuitive because more bike infrastructure usually yields more bike trips (Dill and Carr, 2003; Winters et al., 2016), whereas steeper slopes act as a disincentive for cyclists (Hood et al., 2011; Winters et al., 2016). The coefficient of Linearity was negative, which suggests that more nonlinearity in the bike network is positively associated with BKT. This nonintuitive result can be due to the fact that nonlinear bike links in the City of Vancouver are usually off-street paths. Another explanation would suggest that more nonlinearity means longer trips for the cyclists, which increases the BKT. It is worth mentioning that the impact of network linearity on BKT is potentially different from its impact on bike trip counts, bike mode, and route choices; however, this is not within the scope of the present study.

As for the FB models, the model with spatial effects showed better goodness of fit than the one without spatial effects (i.e., significantly lower DIC). Moreover, the spatial effects proportion of the overall random effects was considerably high ($\psi = 0.99$), which indicates that there is a spatial correlation between the BKT random effects in the adjacent zones. Including the spatial effects reduced the statistical significance of both Cov and WSlope, making them nonsignificant at the 90% level in the spatial effects models.

**Table 2**
Network indicators model 1.

| Linear Regression Model | | | FB Model | | | | FB Spatial Effects Model | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Parameter Estimate | Pr > \|t\| | Estimate | SD | Bayesian Confidence Interval | | Estimate | SD | Bayesian Confidence Interval | |
| | | | | | 2.5% | 97.5% | | | 2.5% | 97.5% |
| Intercept | −2.04 | 0.0002 | 1.048 | 0.12 | 0.75 | 1.34 | −1.596 | 0.54 | −2.67 | −0.53 |
| L | 0.389 | <0.0001 | 0.389 | 0.061 | 0.268 | 0.511 | 0.43 | 0.065 | 0.30 | 0.56 |
| Cov | 2.042 | 0.0146 | 2.044 | 0.829 | 0.419 | 3.676 | 0.99* | 0.92 | −0.87 | 2.74 |
| AvgEdLen | 8.307 | 0.0059 | 8.25 | 2.979 | 2.389 | 14.12 | 6.65 | 2.89 | 0.96 | 12.37 |
| R-Squared | 0.32 | | | | | | | | | |
| Adj R-Squared | 0.30 | | | | | | | | | |
| DIC | | | 536 | | | | 499 | | | |
| $\sigma_u^2$ | | | 0.004 | 0.023 | 0.00027 | 0.025 | 0.003 | 0.00047 | 0.0024 | 0.0042 |
| $\sigma_s^2$ | | | | | | | 0.89 | 0.49 | 0.28 | 2.12 |
| $\psi$ | | | | | | | 0.99 | | | |

All other variables were significant at 95% level or higher.
* Nonsignificant at 90% level.

**Table 3**
Network indicators model 2.

| Linear Regression Model | | | FB Model | | | | FB Spatial Effects Model | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Parameter Estimate | Pr > \|t\| | Estimate | SD | Bayesian Confidence Interval | | Estimate | SD | Bayesian Confidence Interval | |
| | | | | | 2.5% | 97.5% | | | 2.5% | 97.5% |
| Intercept | 11.36 | 0.013 | 1.04 | 0.15 | 0.75 | 1.35 | 7.003 | 4.16 | −1.15 | 15.17 |
| L | 0.403 | <0.0001 | 0.40 | 0.06 | 0.28 | 0.52 | 0.46 | 0.061 | 0.34 | 0.58 |
| WSlope | −0.29** | 0.091 | −0.29** | 0.17 | −0.62 | 0.049 | −0.23* | 0.15 | −0.54 | 0.07 |
| Lin | −11.14 | 0.015 | −10.91 | 4.55 | 1.91 | 19.83 | −7.03** | 4.15 | −1.09 | 15.18 |
| R-Squared | 0.30 | | | | | | | | | |
| Adj R-Squared | 0.28 | | | | | | | | | |
| DIC | | | 540.52 | | | | 501.97 | | | |
| $\sigma_u^2$ | | | 0.003 | 0.008 | 0.00027 | 0.016 | 0.003 | 0.00047 | 0.0024 | 0.0042 |
| $\sigma_s^2$ | | | | | | | 0.87 | 0.45 | 0.33 | 1.99 |
| $\psi$ | | | | | | | 0.99 | | | |

All other variables were significant at 95% level or higher.
* Nonsignificant at 90% level.
** Significant at 90% level.

### 5.2. Land use models

Table 4 shows the parameter estimates of the land use models. Two variables were found significant: total area of the recreational zonings and total area of the residential zonings. The total area of the commercial zonings was found to be non-significantly associated (positive association) with BKT, so it was omitted from the model. The linear regression model showed good fit with an adjusted R-squared equal to 0.49. The results are plausible as Recreational Areas were found to be highly positively associated with BKT, whereas Residential Areas were found to be negatively associated with BKT. Several studies showed that recreational areas would motivate commuters to undertake more trips by bike (e.g., Daley and Rissel, 2011), which was strongly affirmed by the study in hand. Conversely, residential areas can result in lower cycling levels, particularly if there is no specialized infrastructure that can support uninterrupted cycling (Kerr et al., 2016) or if there is a sheer dominance of automobiles in those areas (Kim and Yamashita, 2002) that would make roads less friendly and less safe for cyclists, particularly if there are no traffic calming policies applied (Pucher and Dijkstra, 2003; Pucher and Buehler, 2008).

Moreover, for this category of models, spatial effects model was found to better fit the data (significantly lower DIC), and the spatial effects proportion of the overall random effects was considerably high ($\psi = 0.99$).

It may be argued that the recreational and residential areas in the land use model 1 are acting as proxies for other factors because they were included as absolute values. For example, recreational area may be considered as a proxy for less traffic in recreational areas, whereas residential area may be considered as a proxy for suburban areas with poorer bike infrastructure. Therefore, to confirm the associations between residential and recreational areas and BKT, another model was developed as shown in Table 5, where the relative areas were used instead of the absolute areas (i.e., RecDen is the ratio between recreational areas and the total zone area, whereas ResDen is the ratio between residential area and the total zone area). Although it showed lower R-squared, Land use model 2 confirmed the association results of Land use model 1.

**Table 4**
Land use model 1.

| Linear Regression Model | | | FB Model | | | | FB Spatial Effects Model | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Parameter Estimate | Pr > \|t\| | Estimate | SD | Bayesian Confidence Interval | | Estimate | SD | Bayesian Confidence Interval | |
| | | | | | 2.5% | 97.5% | | | 2.5% | 97.5% |
| Intercept | 1.10 | <0.001 | 1.048 | 0.13 | 0.79 | 1.306 | 1.08 | 0.21 | 0.64 | 1.51 |
| Rec | 3.78 | <0.001 | 3.78 | 0.34 | 3.11 | 4.45 | 3.63 | 0.33 | 2.98 | 4.30 |
| Res | −1.56 | <0.001 | −1.56 | 0.40 | −2.34 | −0.77 | −1.41 | 0.51 | −2.42 | −0.39 |
| R-Squared | 0.498 | | | | | | | | | |
| Adj R-Squared | 0.490 | | | | | | | | | |
| DIC | | | 495 | | | | 471 | | | |
| $\sigma_u^2$ | | | 0.020 | 0.15 | 0.00026 | 0.06 | 0.003 | 0.00047 | 0.0024 | 0.004 |
| $\sigma_s^2$ | | | | | | | 0.53 | 0.27 | 0.25 | 0.61 |
| $\psi$ | | | | | | | 0.99 | | | |

All variables were significant at 95% level or higher.

**Table 5**
Land use model 2.

| Linear Regression Model | | | FB Model | | | | FB Spatial Effects Model | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Parameter Estimate | Pr > \|t\| | Estimate | SD | Bayesian Confidence Interval | | Estimate | SD | Bayesian Confidence Interval | |
| | | | | | 2.5% | 97.5% | | | 2.5% | 97.5% |
| Intercept | 1.10 | 0.082 | 1.04 | 0.15 | 0.74 | 1.35 | 0.43 | 0.39 | −0.36 | 1.20 |
| Area | 0.52 | 0.037 | 0.52 | 0.25 | 0.03 | 1.02 | 0.70 | 0.26 | 0.19 | 1.23 |
| RecDen | 6.37 | <0.001 | 6.36 | 1.32 | 3.76 | 8.97 | 5.49 | 1.26 | 3.00 | 7.97 |
| ResDen | −1.90 | 0.022 | −1.90 | 0.83 | −3.54 | −0.26 | −1.61** | 0.91 | −3.40 | 0.18 |
| R-Squared | 0.30 | | | | | | | | | |
| Adj R-Squared | 0.28 | | | | | | | | | |
| DIC | | | 540.80 | | | | 471 | | | |
| $\sigma_u^2$ | | | 0.003 | 0.008 | 0.00026 | 0.016 | 0.003 | 0.00047 | 0.0024 | 0.004 |
| $\sigma_s^2$ | | | | | | | 0.65 | 0.29 | 0.28 | 0.77 |
| $\psi$ | | | | | | | 0.99 | | | |

All other variables were significant at 95% level or higher.
** Significant at 90% level.

### 5.3. Road facility model

Table 6 shows the results of the road facility model, where both arterial-collector roads proportion (ArtColl) and the off-street bike lanes proportion (Off_St) are the included variables. Arterial-Collector Proportion is found to have a negative association with BKT, which is intuitive as arterial and collector roads are usually perceived by cyclists as less safe and less friendly (Marshall and Garrick, 2011). However, the Off-Street Proportion variable was highly positively associated with BKT, which agrees with several previous studies that showed that cyclists prefer using the off-street routes over the on-street ones (e.g., Winters and Teschke, 2010; Winters et al., 2011; Larsen and El-Geneidy, 2011). Again, the models in this category showed that spatial effects model has better fit (i.e., lower DIC) than the one that did not account for spatial effects (lower DIC), and the spatial effects proportion of the overall random effects was also found considerably high ($\psi$ = 0.99).

### 5.4. Combined models

Finally, various models were developed to combine the variables from the different investigated categories, i.e., network, land use, and road facility. The three models with the highest goodness of fit are presented in this section, as shown in Tables 7–9. The first combined model (Model A) included five variables: Length, Weight Slope, Recreational Area, Residential Area, and Arterial-Collector Proportion. The model showed the best goodness of fit among the different tested models (adj R-squared = 0.61, DIC = 449.89). The second combined model (Model B) included four variables: Coverage, Recreational Area, Residential Area, and Arterial-Collector Proportion. It yielded an adj R-squared of 0.51, and DIC of 467. The third model (Model C) also included four variables: Length, Recreational Area Density, Residential Area Density, and Off-Street Proportion. The adj R-Squared for this model was equal to 0.42 and DIC equal to 487.50.

**Table 6**
Road facility model.

| Linear Regression Model | | | FB Model | | | | FB Spatial Effects Model | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Parameter Estimate | Pr > \|t\| | Estimate | SD | Bayesian Confidence Interval | | Estimate | SD | Bayesian Confidence Interval | |
| | | | | | 2.5% | 97.5% | | | 2.5% | 97.5% |
| Intercept | 0.91 | 0.005 | 1.04 | 0.16 | 0.72 | 1.37 | 0.79 | 0.36 | 0.08 | 1.52 |
| Off_St | 5.46 | <0.001 | 5.46 | 0.95 | 3.59 | 7.35 | 5.15 | 0.98 | 3.22 | 7.09 |
| ArtColl | −1.52** | 0.06 | −1.52** | 0.81 | −3.12 | 0.06 | −1.06* | 0.90 | −2.86 | 0.70 |
| R-Squared | 0.20 | | | | | | | | | |
| Adj R-Squared | 0.18 | | | | | | | | | |
| DIC | | | 557.25 | | | | 538 | | | |
| $\sigma_u^2$ | | | 0.004 | 0.01 | 0.00024 | 0.03 | 0.003 | 0.00047 | 0.0024 | 0.004 |
| $\sigma_s^2$ | | | | | | | 0.48 | 0.18 | 0.24 | 1.17 |
| $\psi$ | | | | | | | 0.99 | | | |

All other variables were significant at 95% level or higher.
* Nonsignificant at 90% level.
** Significant at 90% level.

**Table 7**
Combined model A.

| Linear Regression Model | | | FB Model | | | | FB Spatial Effects Model | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Parameter Estimate | Pr > \|t\| | Estimate | SD | Bayesian Confidence Interval | | Estimate | SD | Bayesian Confidence Interval | |
| | | | | | 2.5% | 97.5% | | | 2.5% | 97.5% |
| Intercept | 0.99 | 0.001 | 1.047 | 0.11 | 0.82 | 1.27 | 1.31 | 0.51 | 0.30 | 12.34 |
| L | 0.41 | <0.001 | 0.41 | 0.07 | 0.27 | 0.54 | 0.38 | 0.07 | 0.23 | 0.51 |
| Rec | 2.32 | <0.001 | 2.33 | 0.38 | 1.58 | 2.27 | 2.25 | 0.39 | 1.48 | 3.05 |
| Res | −3.38 | <0.001 | −3.35 | 0.48 | −4.29 | −2.41 | −2.85 | 0.55 | −3.96 | −1.76 |
| WSlope | −0.26 | 0.04 | −0.26 | 0.13 | −0.51 | −0.005 | −0.19* | 0.13 | −0.45 | 0.06 |
| ArtColl | −1.26 | 0.035 | −1.38 | 0.66 | −2.68 | −0.09 | −0.95* | 0.74 | −2.42 | 0.49 |
| R-Squared | 0.62 | | | | | | | | | |
| Adj R-Squared | 0.61 | | | | | | | | | |
| DIC | | | 462.3 | | | | 449.89 | | | |
| $\sigma_u^2$ | | | 0.0062 | 0.03 | 0.00025 | 0.04 | 0.0033 | 0.00048 | 0.0025 | 0.0044 |
| $\sigma_s^2$ | | | | | | | 0.40 | 0.16 | 0.22 | 0.83 |
| $\psi$ | | | | | | | 0.99 | | | |

All other variables were significant at 95% level or higher.
* Nonsignificant at 90% level.

**Table 8**
Combined model B.

| Linear Regression Model | | | FB Model | | | | FB Spatial Effects Model | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Parameter Estimate | Pr > \|t\| | Estimate | SD | Bayesian Confidence Interval | | Estimate | SD | Bayesian Confidence Interval | |
| | | | | | 2.5% | 97.5% | | | 2.5% | 97.5% |
| Intercept | 1.21 | 0.005 | 1.048 | 0.11 | 0.81 | 1.27 | 0.88 | 0.43 | 0.02 | 1.73 |
| Cov | 2.05 | 0.013 | 2.05 | 0.82 | 0.43 | 3.67 | 2.05 | 0.79 | 0.50 | 3.62 |
| Rec | 3.61 | <0.001 | 3.60 | 0.33 | 2.95 | 4.26 | 3.51 | 0.33 | 2.86 | 4.17 |
| Res | −1.78 | <0.001 | −1.77 | 0.46 | −2.69 | −0.85 | −1.50 | 0.52 | −2.52 | −0.46 |
| ArtColl | −2.01 | 0.012 | −2.01 | 0.79 | −3.59 | −0.44 | −1.31* | 0.83 | −2.96 | 0.31 |
| R-Squared | 0.53 | | | | | | | | | |
| Adj R-Squared | 0.51 | | | | | | | | | |
| DIC | | | 489 | | | | 467 | | | |
| $\sigma_u^2$ | | | 0.015 | 0.10 | 0.00027 | 0.06 | 0.0032 | 0.00048 | 0.0024 | 0.0043 |
| $\sigma_s^2$ | | | | | | | 0.51 | 0.24 | 0.23 | 1.14 |
| $\psi$ | | | | | | | 0.99 | | | |

All other variables were significant at 95% level or higher.
* Nonsignificant at 90% level.

**Table 9**
Combined model C.

| Linear Regression Model | | | FB Model | | | | FB Spatial Effects Model | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Parameter Estimate | Pr > \|t\| | Estimate | SD | Bayesian Confidence Interval | | Estimate | SD | Bayesian Confidence Interval | |
| | | | | | 2.5% | 97.5% | | | 2.5% | 97.5% |
| Intercept | −2.18 | 0.53 | 1.049 | 0.13 | 0.77 | 1.32 | −0.19 | 0.39 | −0.97 | 0.56 |
| L | 0.35 | <0.001 | 0.35 | 0.06 | 0.22 | 0.47 | 0.35 | 0.06 | 0.22 | 0.47 |
| RecDen | 3.36 | 0.008 | 3.35 | 1.27 | 0.85 | 5.85 | 3.24 | 1.20 | 0.91 | 5.63 |
| ResDen | −1.62 | 0.031 | −1.62 | 0.75 | −3.09 | −0.14 | −1.42** | 0.85 | −3.10 | 0.26 |
| Off_St | 2.44 | 0.010 | 2.44 | 0.94 | 0.56 | 4.30 | 1.79** | 0.94 | −0.06 | 3.66 |
| R-Squared | 0.44 | | | | | | | | | |
| Adj R-Squared | 0.42 | | | | | | | | | |
| DIC | | | 512.65 | | | | 487.50 | | | |
| $\sigma_u^2$ | | | 0.007 | 0.03 | 0.00026 | 0.046 | 0.0032 | 0.00047 | 0.0024 | 0.0043 |
| $\sigma_s^2$ | | | | | | | 0.52 | 0.18 | 0.25 | 1.23 |
| $\psi$ | | | | | | | 0.99 | | | |

All other variables were significant at 95% level or higher.
** Significant at 90% level.

The associations between the covariates used in the combined models and BKT agree with the results from the former separated models. All the combined models showed better DIC under the spatial effects model than under the model without spatial effects. Moreover, the spatial effects parameter ($\psi$) was significantly high in all the combined models. However, Weight Slope and Arterial-Collector Proportion variables became nonsignificant upon including the spatial effects.

Application of any of the three combined models for predicting BKT outside the City of Vancouver would be valid only if the models' transferability is tested, particularly when transferred to regions with substantially different levels of cycling or cycling infrastructure. Although model A attained the highest BKT predictability in the City of Vancouver, it contained more constraints than the other models, which would limit its transferability to other regions.

## 6. Conclusion and recommendations

This paper studied the impact of bike network, land use, and road facility on BKT at a zonal level. Models were developed using data from 134 TAZs in the City of Vancouver, Canada. Linear regression and full Bayesian models with and without spatial effects were used as modeling techniques. The developed models showed relatively good fit and revealed important findings. Three bike network indicators (i.e., coverage, continuity, and length) were found to be positively associated with BKT, contrary to the network's slope and linearity indicators that were found to be negatively associated with BKT. In addition, recreational areas were found to be positively associated with BKT, in contrast to residential areas that were negatively associated with BKT. Moreover, arterial-collector roads proportion was shown to be negatively associated with BKT, whereas the proportion of off-street bike paths was positively associated with BKT. The ridership models that accounted for the spatial effects showed better fit. This finding was supported by the significant value of the spatial variation proportion, which would indicate the existence of spatial correlation between the BKT in the adjacent zones. This indicates the importance of accounting for spatial effects when developing cycling ridership models at a macro level as ignoring them could lead to biased results. This can also be an interesting future research area focusing on gaining better understanding of how the value of BKT in a zone affects BKT from adjacent zones. In addition, further research is needed on proximity effects in modeling intra-urban BKT at various levels of aggregation and on other spatial techniques of analyzing BKT such as hotspot analysis and network-based pattern analysis.

These findings support past research suggesting that higher quality bike networks attract more riders. These models show that within a single city, the more well-connected, continuous, flat, dense, recreational, and off-street zones of the bike network are more heavily used, even after accounting for spatial correlation. Higher cycling levels in these zones likely reflect the combined effects of increased bicycle trips (mode choice is influenced by infrastructure quality) and of bicyclist preferences for routes using infrastructure in these zones. As noted in the Introduction section, the BKT models cannot distinguish between induced and diverted travel. There was high variability in BKT among TAZs (coefficient of variability, i.e., standard deviation divided by mean, is approximately two), and more than half of that variance was explained by models using network, land use, and facility variables. Thus, good estimates of BKT at the TAZ level can be generated by these relatively parsimonious models. Zonal BKT estimates can be used in safety and health studies such as calculation of exposure to crashes, physical activity, and air pollution. Furthermore, zonal BKT models can be beneficial in strategic planning applications to identify deficient areas of a city in terms of bike network quality using the predictive network indicators.

Several areas of future research can be investigated based on the present study. First, additional variables can be integrated in the ridership models to investigate further associations with BKT, including built environment variables (e.g., bus stops, traffic signals, and light poles, etc.), socio-demographic variables, additional facility characteristics (buffer type, on-street parking, access points (driveways), lighting, surface type, intersection treatments, etc.), seasonality and time components, street-related factors (e.g., speed limits, level of traffic stress, etc.), and additional bike network indicators that are not correlated to the ones in the study such as betweenness centrality. Moreover, normalizing BKT into bikeability scores can be beneficial for ranking different neighborhoods according to their friendliness to cyclists. The relationship between bikeability and bike network indicators can be a novel area of research. Finally, studying the transferability of the models by applying them to other regions can be useful for validating the developed models and evaluating their accuracy.

## References

Aguero-Valverde, Jonathan, Jovanis, Paul, 2008. Analysis of road crash frequency with spatial models. Transp. Res. Rec.: J. Transp. Res. Board 2061, 55–63.
Buehler, Ralph, Dill, Jennifer, 2016. Bikeway networks: a review of effects on cycling. Transp. Rev. 36 (1), 9–27.
Buehler, Ralph, Pucher, John, 2012. Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes. Transportation 39 (2), 409–432.
Berrigan, David, Pickle, Linda W., Dill, Jennifer, 2010. Associations between street connectivity and active transportation. Int. J. Health Geogr. 9 (1), 1.
Birk, Mia, Roger Geller, 2006. Bridging the gaps: how quality and quantity of a connected bikeway network correlates with increasing bicycle use. In: Transportation Research Board 85th Annual Meeting. No. 06-0667.
Barnes, Gary, Krizek, Kevin, 2005. Estimating bicycling demand. Transp. Res. Rec.: J. Transp. Res. Board 1939, 45–51.
Caulfield, Brian, Brick, Elaine, Thérèse McCarthy, Orla, 2012. Determining bicycle infrastructure preferences – a case study of Dublin. Transp. Res. Part D: Transp. Environ. 17 (5), 413–417.
Daley, Michelle., Rissel, Chris., 2011. Perspectives and images of cycling as a barrier or facilitator of cycling. Transp. Policy 18 (1), 211–216.
Dill, Jennifer, Voros, Kim, 2007. Factors affecting bicycling demand: Initial survey findings from the Portland, Oregon, region. Transp. Res. Rec.: J. Transp. Res. Board 2031, 9–17.
Dill, Jennifer, Carr, Theresa, 2003. Bicycle commuting and facilities in major US cities: if you build them, commuters will use them. Transp. Res. Rec.: J. Transp. Res. Board 1828, 116–123.

Derrible, Sybil, Kennedy, Christopher, 2009. Network analysis of world subway systems using updated graph theory. Transp. Res. Rec.: J. Transp. Res. Board 2112, 17–25.
El-Basyouny, Karim, Sayed, Tarek, 2009. Urban arterial accident prediction models with spatial effects. Transp. Res. Rec.: J. Transp. Res. Board 2102, 27–33.
El Esawey, Mohamed, Lim, Clark, Sayed, Tarek, 2015. Development of a cycling data model: City of Vancouver case study. Can. J. Civ. Eng. 42 (12), 1000–1010.
Fagnant, Daniel J., Kockelman, Kara, 2015. A direct-demand model for bicycle counts: the impacts of level of service and other factors. Environ. Plann. B: Plann. Des. (0265813515602568)
Griswold, Julia, Medury, Aditya, Schneider, Robert, 2011. Pilot models for estimating bicycle intersection volumes. Transp. Res. Rec. 2247, 1–7.
Garrison, William L., Marble, Duane F., 1962. The Structure of Transportation Networks. No. TR62 11. Northwestern Univ Evanston Ill.
Handy, Susan L., Xing, Yan, 2011. Factors correlated with bicycle commuting: a study in six small US cities. Int. J. Sustain. Transp. 5 (2), 91–110.
Haynes, M., Andrzejewski, S., 2010. GIS based bicycle & pedestrian demand forecasting techniques. TMIP Webinar.
Hood, Jeffrey, Sall, Elizabeth, Charlton, Billy, 2011. A GPS-based bicycle route choice model for San Francisco, California. Transp. Lett. 3 (1), 63–75.
Karim, M., Wahba, Mohamed, Sayed, Tarek, 2013. Spatial effects on zone-level collision prediction models. Transp. Res. Rec.: J. Transp. Res. Board 2398, 50–59.
Klobucar, Michael, Fricker, Jon, 2007. Network evaluation tool to improve real and perceived bicycle safety. Transp. Res. Rec.: J. Transp. Res. Board 2031, 25–33.
Kansky, Karel Joseph, 1963. Structure of Transportation Networks: Relationships between Network Geometry and Regional Characteristics.
Kerr, Jacqueline et al, 2016. Perceived neighborhood environmental attributes associated with walking and cycling for transport among adult residents of 17 cities in 12 countries: the IPEN study. Environ. Health Perspect. 124 (3), 290.
Kim, Karl, Yamashita, Eric, 2002. Motor vehicle crashes and land use: empirical analysis from Hawaii. Transp. Res. Rec.: J. Transp. Res. Board 1784, 73–79.
Krizek, Kevin J., Roland, Rio W., 2005. What is at the end of the road? Understanding discontinuities of on-street bicycle lanes in urban settings. Transp. Res. Part D: Transp. Environ. 10 (1), 55–68.
Larsen, Jacob, El-Geneidy, Ahmed, 2011. A travel behavior analysis of urban cycling facilities in Montréal, Canada. Transp. Res. Part D: Transp. Environ. 16 (2), 172–177.
Marshall, Wesley E., Garrick, Norman W., 2011. Evidence on why bike-friendly cities are safer for all road users. Environ. Pract. 13 (01), 16–27.
Mekuria, Mazza C., Furth, Peter G., Nixon, Hilary, 2012. Low-Stress Bicycling and Network Connectivity.
Nelson, Arthur, Allen, David, 1997. If you build them, commuters will use them: association between bicycle facilities and bicycle commuting. Transp. Res. Rec.: J. Transp. Res. Board 1578, 79–83.
Osama, Ahmed, Sayed, Tarek, 2016. Evaluating the impact of bike network indicators on cyclist safety using macro-level collision prediction models. Accid. Anal. Prev. 97, 28–37.
Parkin, John, Wardman, Mark, Page, Matthew, 2008. Estimation of the determinants of bicycle mode share for the journey to work using census data. Transportation 35 (1), 93–109.
Pucher, John, Dijkstra, Lewis, 2003. Promoting safe walking and cycling to improve public health: lessons from the Netherlands and Germany. Am. J. Public Health 93 (9), 1509–1516.
Pucher, John, Buehler, Ralph, 2008. Making cycling irresistible: lessons from the Netherlands, Denmark and Germany. Transp. Rev. 28 (4), 495–528.
Quintero, Liliana, Sayed, Tarek, Wahba, Mohamed M., 2013. Safety models incorporating graph theory based transit indicators. Accid. Anal. Prev. 50, 635–644.
Quintero-Cano, L., Wahba, M., Sayed, T., 2014. Bus networks as graphs: new connectivity indicators with operational characteristics. Can. J. Civ. Eng. 41, 788–799.
Rietveld, Piet, Daniel, Vanessa, 2004. Determinants of bicycle use: do municipal policies matter? Transp. Res. Part A: Policy Pract. 38 (7), 531–550.
Rodrigue, J.P., Comtois, C., Slack, B., 2009. The Geography of Transport Systems. Routledge, New York.
Ryan, G.S., Donlan, J., Ledbetter, L., Arnold, L., Ragland, D., 2010. Seamless Travel: Measuring Bike and Pedestrian Activity in San Diego County and Its Relationship to Land Use, Transportation, Safety, and Facility Type. Alta Planning and Design and Safe Transportation Research and Education Center, University of California, Berkeley.
Schoner, Jessica E., Levinson, David M., 2014. The missing link: bicycle infrastructure networks and ridership in 74 US cities. Transportation 41 (6), 1187–1204.
Sanders, Rebecca L., 2014. Roadway design preferences among drivers and bicyclists in the bay area. In: TRB 93rd Annual Meeting Compendium of Papers. No. 14-5454.
Spiegelhalter, David J. et al, 2002. Bayesian measures of model complexity and fit. J. Roy. Stat. Soc.: Ser. B (Stat. Methodol.) 64 (4), 583–639.
Strauss, Jillian, Miranda-Moreno, Luis F., 2013. Spatial modeling of bicycle activity at signalized intersections. J. Transp. Land Use 6 (2), 47–58.
Schneider, Robert J., Stefanich, Joseph, 2015. Neighborhood characteristics that support bicycle commuting: analysis of the top 100 US census tracts. In: Transportation Research Board 94th Annual Meeting. No. 15-3320.
Scheltema, E.B., 2012. Manual ReCYCLE City: Strengthening the Bikeability from Home to the Dutch Railway Station. Delft University of Technology, Diss. TU Delft.
Tilahun, Nebiyou Y., Levinson, David M., Krizek, Kevin J., 2007. Trails, lanes, or traffic: valuing bicycle facilities with an adaptive stated preference survey. Transp. Res. Part A: Policy Pract. 41 (4), 287–301.
Wilkinson, William C., 1994. Selecting Roadway Design Treatments to Accommodate Bicycles.
Winters, Meghan et al, 2016. Bike Score®: associations between urban bikeability and cycling behavior in 24 cities. Int. J. Behav. Nutr. Phys. Activity 13 (1), 1.
Winters, Meghan, Teschke, Kay, 2010. Route preferences among adults in the near market for bicycling: findings of the cycling in cities study. Am. J. Health Promot. 25 (1), 40–47.
Winters, Meghan. et al, 2011. Motivators and deterrents of bicycling: comparing influences on decisions to ride. Transportation 38 (1), 153–168.
Xie, Feng, Levinson, David, 2007. Measuring the structure of road networks. Geogr. Anal. 39 (3), 336–356.
Yigitcanlar, Tan, Fatih, Dur, 2010. Developing a sustainability assessment model: the sustainable infrastructure, land-use, environment and transport model. Sustainability 2 (1), 321–340.