How Do Land Use, Built Environment and Transportation Facilities Affect Bike-Sharing Trip Destinations?

Ahmed JABER¹, Bálint CSONKA²

1 Corresponding author, ahjaber6@edu.bme.hu, Department of Transport Technology and Economics, Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics

2 csonka.balint@kjk.bme.hu, Department of Transport Technology and Economics, Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics

ABSTRACT

The purpose of this research is to investigate the effect of land use, built environment and public transportation facilities’ locations on destinations of bike-sharing trips in an urban setting. Several methods have been applied to determine the relationship between predicting variables and trip destinations, such as ordinary least squares regression, spatial regression and geographically weighted regression. Additionally, a comparison between the proposed models, count models and random forest has been conducted. The data were collected in Budapest, Hungary. It has been found that touristic points of interest, and healthcare and educational points have a positive impact on bike-sharing destinations. Public transportation stops for buses, trains and trams attract bike-sharing users, which has a potential for the bike-and-ride system. Land use has different effects on bike-sharing trip destinations; mostly as a circular shape variation within the urban structure of the city, such as residential, industrial, commercial and educational zones. Other variables, such as road length and water areas, form as constraints to bike-sharing trip destinations. Geographically weighted and spatial regression performs better than count models and random forest. This study helps decision-makers in predicting the origin-destination matrix of bike-sharing trips based on the transportation network and land use.

KEYWORDS

bike-sharing; public transportation; land use; spatial analysis; trip destination.

1. INTRODUCTION

Cycling promotion is a strategy for reducing traffic congestion and increasing transportation sustainability through lower emissions and land use. Cycling also increases physical activity and improves overall public health. This is why cities are introducing cycling programs: to help achieve goals related to public health, sustainability and the quality of life [1, 2].

Countries are now shifting towards more sustainable and environmentally-friendly transportation solutions that provide pollution-free transportation modes, e.g. cycling. The popularity of cycling is on the rise, and countries compete in developing and integrating this transportation mode into their systems. Concerns about energy efficiency, climate change and transportation congestion have persisted over the last few decades. The bike-sharing concept, which is just one of the many cycling programs, has been increasingly popular in recent years [3]. Bike-sharing systems (BSS) are beneficial to public health since they are inexpensive to operate,
take up little space and promote physical activity, which is especially significant in developing areas [4]. Spatial dispersion of built environment and land use directly influence the BSS and how it should be integrated [5].

There are a few ways to describe bike-sharing systems, but the two most common are docked and dockless [6]. Docked bike sharing allows users to rent and return bicycles at predetermined docking locations. Riders can use mobile phone applications (apps) to rent and return bikes in designated physical or electric-fencing areas for the dockless bike-sharing system [7, 8]. These two bike-sharing systems help promote urban mobility.

This study measures the spatial behaviour of bike-sharing users considering road and railway network characteristics, transport facilities and land use classification. The data were collected in Budapest City, Hungary. Ordinary least squares, spatial autoregressive and geographic weighted regression analyses were generated. These techniques were used to investigate the connections between bike-sharing trip destinations, land use and public transportation stop locations. Specifically, the trip distribution of bike-share users was investigated using a multilevel modelling technique. The objective is to create a statistical model that can be used by transportation planners and city officials to make predictions about how many people will use bicycle sharing and where they choose to utilise it. As a result, the following are the main contributions of this work: (a) geographical analysis of the zonal distribution of bike-sharing trip destinations; and (b) evaluation and comparisons of spatial regressions, geographically weighted regression, ordinary least squares, count models and tree-decision models.

This is the outline of the paper: the literature review for the application of these methods in the bike-sharing prediction that described the relationship between bikes with land use and the road/railway network is given in the next section. The methodologies are described in further depth in Section 3, which is followed by the findings and a discussion of those findings in Section 4. At the end, conclusions are drawn and recommendations for moving forward are presented.

2. LITERATURE REVIEW

2.1 Overview

A number of studies have been carried out to foretell the future of BSS usage [9–11] and investigate the effects of land use and the built environment that are disregarded in favour of temporal and meteorological variables in a time series analysis. For example, to model bicycle demand, Faghih-Imani and Elurub [9] used information from New York City’s bike-sharing system while accounting for spatial and temporal effects. To that end, BSS infrastructure and temporal attributes were used to estimate spatial lag and spatial error models. Arrival and departure rates at BSS stations were found to be spatially and temporally dependent. The enhanced spatial and temporal accuracy of the models is demonstrated by a hold-out sample validation exercise. More models, such as ordinary least squares and geographically weighted regression, were developed in our study, with a focus on spatial attributes of land use that were not studied in previous research.

Demand models for a proposed city expansion were estimated and evaluated with the use of an electric bike-sharing system [12]. Factors such as jobs, demography, bars, restaurants and closeness to a central location were identified as significant predictors of variation explained. Furthermore, because it takes into account spatial dependence, spatial regression outperforms random forest. Our study, on the other hand, makes a contribution by looking into more spatial factors, such as educational, healthcare and other facilities.

Using annual trip data, the ridership of Toronto’s bike-sharing system was studied to determine what factors contribute to the city’s high usage rate [5]. A thorough spatial analysis revealed the effects of socio-demographic factors, universities, public transportation, bike infrastructure and weather on bike share usage. There was a correlation between road network topology (intersection density and station spatial dispersion) and the demand for bike sharing, according to empirical models. Bicycle infrastructure (bike lanes, paths
and so on) also contributes to increased demand for bike-sharing. The influence of points of interest, socio-economic factors and transit stations on bike-sharing demand in Montreal, Canada, was also studied, using a spatially variable regression model [13]. The suggested approach outperformed both classical machine learning and geo-statistical methods in terms of out-of-sample prediction accuracy. Rixey [14] analysed the average monthly demand for bike sharing in three US cities and compared the impact of common characteristics. Faghih-Imania et al. [15] assessed Montreal’s hourly demand for bike sharing, accounting for both the weather and the time of day. More studies are being undertaken using this approach to analyse how climate, time of year and population distribution of different cities affect the popularity of bike-sharing programs [16–18]. In Shen et al. study [19], for nine days in a row, authors utilised a spatial autoregressive model to examine what characteristics affect people’s propensity to use dockless bikes (SAR). The introduction of spatial impact was accomplished through the use of a spatial weight matrix where the nodes are the inverse distance between neighbourhoods within 5 kilometres.

The main differences between our research and the previous studies are: touristic points of interest (POI), crossings, traffic signals, transit stops, educational, healthcare facilities, waterways, road length, green areas, water areas, cemetery areas, industrial areas, commercial areas, military areas, recreational areas and residential areas were not considered, while our study has involved all of these factors.

Additionally, traditional geographically weighted regression (GWR) was employed to capture the geographic variability of the impact of POI on bike-sharing demand by [20]. They uncovered a gap across station types in terms of the influence of variables including bicycle infrastructure, station capacity and socioeconomic indicators. In a related study, Munira and Sener [21] used geographically weighted Poisson regression to examine the effects of socioeconomic status and school on bicycle use. The spatially changing relationship in Chicago was investigated by [22, 23] using GWR. However, in our research, we looked at how GWR stacked up against other models.

Moreover, the built environment significantly influenced the amount of bike-sharing reallocations in Nanjing; for instance, docked stations in areas with a higher concentration of restaurants and service sector jobs required more morning bike removal and evening bike filling [24]. Station-level bike-sharing utilisation was measured in connection to demographic and socioeconomic variables and land use and transportation infrastructure using Bayesian regression [17]. Xu et al. [25] found a correlation between the population density, commercial density and the number of crossings in Singapore and the need for bike sharing there.

### 2.2 Analytical approach

As far as modelling goes, most prior research on the topic of how built environment affects bike ridership relied on standard regression models. Several studies have made an effort over the past few years to tackle the autocorrelation problem. For instance, [9], [19] and [26] utilised spatial error and spatial lag models as two types of spatial regression to analyse the impact of location and time on cycling. To analyse the geographical connection between each pair of stations, a spatial multiple regression model was used. Global spatial regression models, in which the coefficient for each variable is the same throughout, were commonly utilised in these studies. Thus, they neglected to account for the influence of geographical heterogeneity. For this reason, GWR, as a local-based regression model, is a useful tool for addressing the issue of spatial non-stationarity. The geographically weighted regression is also commonly used in the transport sector to investigate the local impacts of related factors on ridership [27], public transportation [28–30], parking demand [31], traffic [32] and accidents [33]. They all concurred that this significantly outperforms global regression models on the basis of goodness of fit. However, this model has only been used in a limited research to analyse the spatial relationship between land use, built environment and bike-sharing usage, and no research has developed this method to identify land use classifications and road/railway facilities in order to investigate the impact on bike-sharing destinations. This research aims to fill that gap by examining how
these factors affect where people end up when they take a bike-share. In our research, we looked into, compared and researched each of these theories to produce substantial findings.

3. METHODOLOGY

3.1 Data investigation

Budapest has about 1.7 million inhabitants and covers a total area of 525 km² consisting of 23 districts. Universities and colleges attract over 1 million students [34]. Due to its characteristics, Budapest has local and inter-city traffic. Its public transportation includes 4 metro lines, 260 bus lines and 30 tram routes serving roughly five million trips every day [35]. Budapest has the fifth highest public transport modal share (45%) among the European cities with a population of over one million [36]. Cycling’s share has risen in recent years, reaching 23% in 2015 [37]. The city promotes intermodal mobility options, with more than ten bike-sharing companies [38] conducting over two thousand trips per day [39].

Data about bike-sharing trips, road infrastructure, land use and built environment were collected. Bike-sharing trip data for the period between 12 and 25 September 2021 were provided by The Donkey Republic bike-sharing company in Budapest. Trip information included station ID, trip start and end points, station latitude and longitude and station capacity. After eliminating duplicates and other errors, we are left with 47,630 journeys (e.g. without complete trip information and trip durations shorter than 30 s). The independent components of road infrastructure, land use and built environment are the locations of public transit stops, traffic signals, road crossings, railways network, road network, waterways and land use category. The following land use categories are distinguished in Budapest: green, commercial, industrial, residential, water, cemetery, healthcare, recreational and military areas, locations of educational and healthcare facilities, touristic points of interest and facilities which can be obtained through the Open Street Map data. Touristic POI data includes the latitude and longitude of the categories.

Furthermore, a grid-cell based technique was used to disclose the spatial distribution of travel destinations in terms of density per zone. Each cell is 1 km² in size. This grid is recommended in [40, 41] and provides a smaller area with a higher number of zones (531) than the number of districts (23) and the number of sub-districts (203), giving more precision.

It is important to test whether there are correlations between the predicting variables for modelling reliability and prediction accuracy. This step could be done by different ways, such as: the variance inflation factor, Spearman’s correlation, etc. A Spearman’s correlation matrix was built to avoid multi-collinearity in order to reveal the correlation between independent variables and target variables [22, 42], as shown in Figure 1. Each variable with a greater correlation ($\rho > 0.6$) to the total number of bike-sharing trip destinations was removed from the analysis. The correlation coefficient matrix shows that the facilities density, crossings density and bicycle lots are highly correlated with other variables. Similarly, waterways and water area are correlated. Green areas and other land use classifications are correlated as well. Built-up area is correlated with most of the variables. Finally, bike station density and their capacity are highly correlated. The final considered independent variables are touristic points, traffic signals, bus stops, rail/tram stops, healthcare points, educational points, rail/tram length, roads length, bike stations capacity, water area, cemetery area, commercial area, military area, industrial area, recreational area and residential area.

In summary, all the predicting variables are described in Table 1, showing the descriptive statistics of predicting variables per zone (the mean and the standard error). It is observed that, on average, roads cover 19.77 km per zone, while railways cover 3.71 km per zone. For the different geographical points, there are close to 8 bus stops, 2 rail/tram stops, 1 educational facility, 2 healthcare facilities, 8 touristic POIs and 8 traffic signals. The bike-sharing destination trips per zone amount to 212.63 daily trips. The land use areas are mostly residential and green areas (green areas were eliminated as per the correlation matrix).
To analyse the influence of land use and road network integration with bike-sharing systems, we carried out the following analyses:

- Ordinary least square model,
- Geographic weighted regression,
- Spatial regression.
Each analysis method has its own perspective. The main point of using multiple methods is to predict bike-sharing trip destinations, and then to compare model performance results. For instance, to determine the effects of the predicting variables on the zonal destination of trips, an ordinary least squares model was developed. Using spatial regression and geographic weighted regression, we may better visualise the coefficients in a spatial context and make comparisons across different zones. Independent variables such as density have varying localised effects on bike trips due to the geographical clustering and heterogeneity of built environment elements [43].

We split the database into training and test sets with an 80:20 ratio using the random split. The training set was used in building the spatial and statistical models, while we used the 20% to test these models for validation purposes.

In the following sub-sections, we present the basics and formulas for each model that we used in the analysis.

**Ordinary least square model**

Multiple linear regression, or OLS regression, is a common method for examining the connection between independent and dependent variables. Assumptions important to the model include the independency of individual observations and the stability of the association between predictor and response variables. This method uses this Equation 1 below:

\[
y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i
\]  

(1)

where: \( y_i \) represents the observation \( i \) of the target variable; \( \beta_{ik} \) is the coefficient of the independent variable \( k \) for observation \( i \); \( x_{ik} \) is the \( k \)-th independent variable for the \( i \)-th observation; and \( \epsilon_i \) is the error.

**Spatial auto-regressive model**

Previous research has showed that the spatial autoregressive model is an excellent tool for studying bike-sharing behaviours [19]. Spatial autoregressive models require a weights matrix, and these models can be defined in a variety of ways, including by contingency, distance, and k-nearest neighbours. We did not consider distance-based or k-nearest neighbours-based matrices were the best option because of the wide range in distances between observations in our dataset. Instead, we employed a contingency-based matrix with a queen-type neighbourhood, which means that two observations were deemed neighbours if they shared a common point. This type of weight necessitates the use of polygons rather than points for observations [44].

Either a spatial lag model (SLM) or a spatial error model (SEM) can be chosen based on the results of the Lagrange multiplier test. With the spatial lag model, the dependent variable’s geographically weighted values are included as an explanatory variable to take into account spillover effects. For the formula, see below:

\[
y = \rho W y + \beta X + \epsilon
\]  

(2)

where \( y \) is an \( N \times 1 \) vector of observations on the dependent variable; \( X \) is an \( N \times K \) matrix of observations on the explanatory variables; \( \beta \) represents a \( K \times 1 \) vector of regression slope coefficients; \( \epsilon \) is the error term vector; \( \rho \) is the coefficient of spatial autocorrelation, and \( W \) is the matrix of spatial weights.

In the spatial error model, the error term values are spatially weighted. The formula is given by:

\[
y = \beta X + e
\]  

(3)

where \( e = \lambda W e + \epsilon \), and \( \lambda \) is the autoregressive coefficient.

**Geographic weighted regression**

This model employs a weighted linear regression technique (the weight in this model is determined by the distance between two observations) [13]. The GWR model provides a spatial extension to the standard OLS model, and its function is as follows:
\[ y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{n} \beta_{ik}(u_i, v_i)x_{ik} + \epsilon_i \]  

where: \( y_i \) is the \( i \)-th observation of the target variable; \( \beta_{ik} \) represents the coefficient of the \( k \)-th independent variable; \( x_{ik} \) represents the \( k \)-th independent variable; \( \epsilon_i \) is the error; \((u_i, v_i)\) is the spatial coordinates [45], all of which for the \( i \)-th observations.

In the GWR model, the impact of predictive variables on the response varies by location. At each bike station, an independent variable has a local coefficient. Weights \( w_{ij} \) are assigned to other data points \((j)\) based on their distances to the studied data point \((i)\) when estimating model coefficients at one data point. In order to estimate the coefficients of the independent variables, the GWR model seeks to minimise the weighted sum of squares. The objective function is illustrated below:

\[ \sum_{j=1}^{n} w_{ij}(y_i - \beta_0 - \sum_{k=1}^{p} \beta_{ik}x_{ik})^2 \]  

4. RESULTS AND DISCUSSION

4.1 OLS and spatial regression

We ran an analysis on the database to check for spatial autocorrelation so as to determine if spatial effects needed to be taken into account. The global Moran’s I statistics, which reveal whether variable values tend to form spatial clusters, were computed for the bike-sharing trip destinations. Moran’s I value for all trips amounted to 0.749, p-value < 0.0001, indicating the presence of highly substantial positive spatial autocorrelation [46]. Moran’s I is -1 to +1. Higher positive values indicate that close observations have similar attribute values, while distant observations have distinct attribute values, showing spatial aggregation. A negative value, on the other hand, implies spatial dispersion, and a value close to zero indicates a spatially random distribution. Moran’s I test’s null hypothesis is that independent variables are spatially independent, yielding a low Moran’s I statistic [44]. In addition, we examined OLS, SEM and SLM model performances using the Akaike information criterion (AIC) and log-likelihood statistics, as did [47]. Specifically, AIC is a prominent metric for measuring the goodness-of-fit of a spatial regression model. The final model chosen is with the lowest AIC [48], as shown in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>SEM</th>
<th>SLM</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>703.38</td>
<td>667.25</td>
<td>689.17</td>
</tr>
<tr>
<td>Likelihood</td>
<td>-350.07</td>
<td>-335.82</td>
<td>-340.33</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.637</td>
<td>0.832</td>
<td>0.815</td>
</tr>
</tbody>
</table>

As a result, the spatial error model, which takes spatial autocorrelation into account, was used. Table 3 displays the estimation results of the models.

As shown in Table 3, most of the variables are significant either at 90%, 95% or 99% levels, except cemetery areas, railway length, military areas and recreational areas. The touristic points of interest attract bike sharing trips with a utility of 7.54. This is expected as a large amount of users are visitors who are interested to see the touristic locations of the city. Regarding road and railway networks and facilities, it is found that traffic signals reduce the number of trips to that zone. This indicates that bike-sharing users do not prefer trips to dense traffic locations and signalised intersections. Bike-sharing destination trips are less likely to go to a certain zone if that zone has a longer road network; the denser the roads, the less likely people are to ride bikes there. Again, this is explained by the conflict with other vehicle types, which bikers find less attractive to interact with. For public transportation stops, both bus and rail/trams stops have positive influence on bike-sharing trip destinations. This leads to the conclusion that the city has a potential for a
Figure 2 – Spatial distribution of the regression coefficients estimated for significant variables

- a) Bus stops
- b) Rail/tram stops
- c) Traffic signals
- d) Touristic POI
- e) BSS capacity
- f) Roads length
- g) Water area
- h) Industrial area
- i) Residential area
- j) Commercial area
- k) Educational points
- l) Healthcare points
Table 3 – Results estimation of the predicting variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>79.5771</td>
<td>7.0094</td>
<td>0.06243 *</td>
</tr>
<tr>
<td>Touristic POI</td>
<td>7.54238</td>
<td>1.22752</td>
<td>0.00000 ***</td>
</tr>
<tr>
<td>Traffic Signals</td>
<td>-2.39404</td>
<td>1.82379</td>
<td>0.08929 *</td>
</tr>
<tr>
<td>Educational Points</td>
<td>12.1528</td>
<td>9.3317</td>
<td>0.09281 *</td>
</tr>
<tr>
<td>Capacity of BSS</td>
<td>23.2041</td>
<td>2.08024</td>
<td>0.00000 ***</td>
</tr>
<tr>
<td>Water Area</td>
<td>-2.21322</td>
<td>0.776961</td>
<td>0.00439 ***</td>
</tr>
<tr>
<td>Commercial Area</td>
<td>0.492441</td>
<td>0.88455</td>
<td>0.07772 *</td>
</tr>
<tr>
<td>Healthcare Points</td>
<td>9.47515</td>
<td>4.81466</td>
<td>0.04907 **</td>
</tr>
<tr>
<td>Residential Area</td>
<td>0.567271</td>
<td>0.420797</td>
<td>0.07763 *</td>
</tr>
<tr>
<td>Railway km</td>
<td>-0.149154</td>
<td>1.68998</td>
<td>0.92967</td>
</tr>
<tr>
<td>Road km</td>
<td>-7.62236</td>
<td>1.91854</td>
<td>0.00007 ***</td>
</tr>
<tr>
<td>Industrial Area</td>
<td>0.203833</td>
<td>0.530287</td>
<td>0.07007 *</td>
</tr>
<tr>
<td>Military Area</td>
<td>0.283286</td>
<td>2.4527</td>
<td>0.90805</td>
</tr>
<tr>
<td>Recreation Area</td>
<td>1.03886</td>
<td>7.73989</td>
<td>0.89323</td>
</tr>
<tr>
<td>Cemetery Area</td>
<td>-0.62335</td>
<td>1.28403</td>
<td>0.62735</td>
</tr>
<tr>
<td>Bus Stops</td>
<td>3.91237</td>
<td>2.01006</td>
<td>0.04161 **</td>
</tr>
<tr>
<td>Rail/Tram Stops</td>
<td>5.83992</td>
<td>4.30675</td>
<td>0.07510 *</td>
</tr>
<tr>
<td>Lambda</td>
<td>0.888628</td>
<td>0.023573</td>
<td>0.00000 ***</td>
</tr>
</tbody>
</table>

* Significant at 90% level; ** significant at 95% level; *** significant at 99% level.

bike-and-ride system which involves ending bike-sharing trips within zones that have public transportation stops nearby. It is also found rail/tram stops have more impact then bus stop density on bike-sharing stops by a ratio of 6:4 approximately. Additionally, it is noticed that when the capacity of bike-sharing stations increases, bike-sharing trips also increase, with a coefficient of 23.2.

In regards to land use classifications, it is found that residential areas have a positive impact on bike-sharing destinations with a coefficient of 0.57. This is explained by the fact that home-based trips end in residential areas. That points to the need of establishing bike-sharing stations near dense residential areas. Furthermore, bike-sharing trips are significant in commercial areas as well, with a coefficient of 0.49. This is expected, as these areas have a lot of attraction points for all the city population. In the aspects of industrial areas, there is a lower positive impact with a 0.20 coefficient. This could be due to limited accessibility of public transportation at these locations, so workers may use bike-sharing to commute. This needs more future investigation. However, zones with water areas are less attractive to bike-sharing users. Finally, both educational and healthcare facilities have large influence on bike-sharing trips with the coefficients 12.2 and 9.5, respectively. Educational influence is explained by students and teachers who use bike-sharing to access study or work locations. Hospitals and clinics have relatively lower impact than educational facilities. Their influence is explained by the public safety need to not interact with other residents when visiting a healthcare facility.

4.2 GWR

The geographical weighted regression results showed that the model could explain 82.3% of the spatial variation in bike-sharing destination trips with AIC equalling 694.88. We used the same predicting variables used in the spatial error model based on the training set. The geographical results predict the variation of each variable and its impact on bike-sharing trip destinations. It is found that the results of GWR and spatial error model are somehow similar and better than OLS. The degree of variation of each variable showed different characteristics, as shown in Figure 2.
We have visualised the spatial distributions for coefficients for the significant variables as shown in Figure 2, where red colour represents the highest and blue the lowest values. For bus stops as shown in Figure 2a, the positive values are in the central and northern-western parts of the city, while the negative values are around the southern-eastern sides of the city. It can be seen that the positive values are in commercial, residential and touristic districts, while the negative ones are in the countryside and main industrial hubs. As for rail/tram stops, it is clearly visible that the concentration is within the centre and surrounding zones that include the main railway stations in the city. The shape seems to be circular regarding the railway network. As for traffic signals, the shape is circular as well, with a shifting towards the western side of the city, including the centre, where traffic movement is larger. For touristic points of interest, it is observed that the concentration is within the centre and in the eastern and southern sides of the city. This area contains several parks that attract visitors. For bike-sharing station capacity, it is necessary to note that all the coefficients are positive, but with higher values in the south. The road length coefficients are clearly opposite from the previous variables. In the city centre, the coefficients are negative, which means there are less bike-sharing destination trips. This variable stands as a constraint for the increase of bike-sharing trips within urban areas. The more freedom in movement given to bikers, the more they prefer using the network.

In the aspects of land use classifications, it is observed that industrial, residential, commercial and educational areas and healthcare points act in similar ways, with some differences as described below:
- Residential area: with a fully circular shape, forming the structure of housing distribution in the city.
- Industrial area: the coefficients shift towards the northern-western sides of the city.
- Commercial area: there is a clear shift to the eastern side of the Danube river, including the city centre where most of the shopping malls are concentrated.
- Healthcare points: they act in a similar way to commercial areas, as per the concentration of the hospitals.
- Educational points: the shifting here is to the north of the city, with an additional area in the south. Research shows that these areas have a greater number of schools and universities.
- Water area: it acts in an opposite way to other land use classifications, forming another constraint to attracting bike-sharing trips.

4.3 Validation

The potential to anticipate probable bike-sharing trip destinations is crucial in establishing the goodness of a regression model. One further way to tell if a regression is over-fitted is to look at the splitting prediction [13]. Therefore, in the spatial regression we evaluated the two samples prediction of the proposed model by reserving randomly 20% trips as stated in the methodology. As a result, we use the root-mean-squared error (RMSE) and R-squared sets to evaluate the proposed models. We have found that the R-squared and RMSE differences between the training and test sets are relatively close, as shown in Table 4.

<table>
<thead>
<tr>
<th>Model</th>
<th>R-squared training set</th>
<th>R-squared test set</th>
<th>RMSE training set</th>
<th>RMSE test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEM</td>
<td>0.832</td>
<td>0.809</td>
<td>123.6</td>
<td>126.6</td>
</tr>
</tbody>
</table>

For the geographically weighted regression, the main validation method is the bandwidth, which depends on either using AIC, cross validation CV or bandwidth parameter methods [13, 33, 43, 49]. The bandwidth parameter method is used to implement the GWR model. It is accomplished primarily through an incremental bandwidth selection approach of selecting the best bandwidth to achieve an optimal GWR model. As a result, the software runs the GWR multiple times with varying bandwidths in order to find the optimal one. GWR residuals should be randomly distributed, so the optimal distance for GWR analysis is chosen. As a result, GWR models with a distance of 4967 m outperformed the other models; this value is a reasonable distance at which land use, built environment and transportation facilities may have the greatest influence on bike-sharing trip destinations. Thus, a bandwidth of 4967 m was chosen as the optimal GWR model.
Furthermore, we compared the proposed models to the reference models in terms of the performance, such as:
- Random forest: randomly construct and merge a multitude of decision trees to predict the demand.
- Poisson regression
- Negative binomial regression

The prediction results of the proposed method and benchmark models are shown in Table 5. Overall, geo-statistical models outperform random forest and count models, demonstrating the value of geographic information in predicting bike-sharing destinations.

<table>
<thead>
<tr>
<th>Model</th>
<th>R-squared</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEM</td>
<td>0.832</td>
<td>123.6</td>
</tr>
<tr>
<td>GWR</td>
<td>0.823</td>
<td>114.0</td>
</tr>
<tr>
<td>OLS</td>
<td>0.637</td>
<td>202.3</td>
</tr>
<tr>
<td>RF</td>
<td>0.674</td>
<td>229.3</td>
</tr>
<tr>
<td>Poisson</td>
<td>0.683</td>
<td>312.9</td>
</tr>
<tr>
<td>NB</td>
<td>0.695</td>
<td>312.4</td>
</tr>
</tbody>
</table>

5. CONCLUSIONS

This research focuses on how various spatial factors influence bike-sharing destinations in an urban setting. Special attention is paid to the interdependencies of road and railway networks, points of interest, land use and bike-sharing. The main novelty of this study is that the effect of land use, railway and road facilities on bike-sharing trip spatial distribution was analysed. The regression analysis revealed a strong relationship between touristic points of interest and capacities of bike-sharing stations and bike-share usage. Commuting to work, healthcare and education appear to account for a sizable portion of bike-sharing trips. We found positive associations between land use classifications and the number of trips in the regression analysis. Concerning transportation aspects, bike-sharing trips are influenced by road length in the zone, public transportation stops, capacity of bike-sharing stations and traffic signals. In the geographically weighted regression, the bike-sharing spatial distribution is affected by most of the variables in a circular way starting from the centre of the city towards the countryside. Some exceptions were investigated in high-concentration educational and healthcare hubs. Other variables formed as constraints, such as road length and water areas. Comparing the different methods of analysis, we observed that geographically weighted regression and spatial error regression are better than linear regression, count models and random forest. The limitations of the study that we have not considered are the terrain differences in the city, which could have an impact on the distribution of bike-sharing trip destinations. Further research may focus on other micro-mobility options in temporal and spatial aspects to contribute to more sustainable transport and improve mobility, as well as consider the contour elevations of the city. Additionally, the e-bike sharing service could be a good motivation to build on in the next studies.

REFERENCES


[34] KSH. Data of population by main characteristics of education by region from Microcensus 2016. [https://statinfo.ksh.hu/Statinfo/haViewer.jsp](https://statinfo.ksh.hu/Statinfo/haViewer.jsp) [Cited Apr. 2022].


[41] Hou H, Estoque RC, Murayama Y. Spatiotemporal analysis of urban growth in three African capital cities: A


Cómo se relacionan los usos de bicicleta y la geografía de la ciudad?

Ahmed Jaber, Universidad Norteamericana

Este estudio busca identificar la relación entre el uso de bicicletas y la geografía de la ciudad. Se han utilizado datos de sensación remota para analizar el uso de bicicletas en diferentes áreas de la ciudad. Se ha encontrado que la ubicación geográfica y la infraestructura urbana son factores clave en el uso de bicicletas. Los resultados indican que la accesibilidad geográfica y la disponibilidad de infraestructura urbana son factores determinantes en el uso de bicicletas en la ciudad. Estos hallazgos pueden ser útiles para la planificación urbanística y la promoción del uso de bicicletas como medio de transporte sostenible.