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

Ahmed JABER[1], Bálint CSONKA[2]

[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 will 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).

	TP	F	Cr	TS	BL	BS	RTS	EP	HP	RTL	RL	W	BA	BiS	BSC	WA	GA	CA	CoA	HA	IA	MA	RcA	RsA
Touristic POI (TP)	X																							
Facilities (F)	0.76	X																						
Crossings (Cr)	0.42	0.57	X																					
Traffic Signals (TS)	0.43	0.59	0.76	X																				
Bicycle Lots (BL)	0.77	0.86	0.55	0.57	X																			
Bus Stops (BS)	0.33	0.47	0.61	0.42	0.40	X																		
Rail/Tram Stops (RTS)	0.34	0.43	0.59	0.58	0.40	0.48	X																	
Educational Points (EP)	0.37	0.53	0.34	0.33	0.40	0.34	0.24	X																
Health Points (HP)	0.33	0.78	0.58	0.57	0.65	0.49	0.43	0.48	X															
Rail/Tram Length (RTL)	0.07	0.13	0.29	0.26	0.12	0.21	0.38	0.07	0.12	X														
Roads Length (RL	0.34	0.42	0.70	0.55	0.44	0.56	0.44	0.25	0.45	0.18	X													
Waterways (W)	-0.03	-0.08	-0.12	-0.06	-0.02	-0.19	-0.06	-0.10	-0.11	-0.05	-0.17	X												
Builtup Area (BA)	0.46	0.68	0.58	0.53	0.55	0.55	0.39	0.47	0.64	0.21	0.50	-0.21	X											
Bike Stations (BiS)	0.45	0.70	0.42	0.41	0.66	0.29	0.29	0.37	0.53	0.09	0.24	-0.04	0.50	X										
BS Capacity (BSC)	0.45	0.69	0.40	0.41	0.66	0.29	0.29	0.34	0.51	0.08	0.23	-0.02	0.47	0.97	X									
Water Area (WA)	0.11	0.02	-0.02	0.07	0.11	-0.14	0.03	-0.07	-0.03	-0.02	-0.16	0.61	-0.12	0.05	0.08	X								
Green Area (GA)	-0.10	-0.23	-0.35	-0.24	-0.17	-0.48	-0.25	-0.22	-0.28	-0.21	-0.40	0.16	-0.57	-0.13	-0.12	-0.05	X							
Cemetery Area (CA)	-0.01	-0.04	-0.05	-0.03	-0.04	-0.05	0.02	-0.06	-0.05	0.07	0.08	-0.06	-0.07	-0.03	-0.03	-0.05	-0.09	X						
Commercial Area (CoA)	-0.03	0.04	0.15	0.15	-0.01	0.05	0.17	-0.03	0.05	0.17	0.18	0.00	0.14	-0.03	-0.03	-0.04	-0.06	-0.02	X					
Health Area (HA)	-0.03	-0.03	-0.05	-0.02	-0.03	-0.02	0.02	-0.04	0.33	0.02	-0.01	0.13	-0.06	-0.03	-0.03	0.03	0.03	-0.02	0.00	X				
Industrial Area (IA)	-0.14	-0.14	0.00	-0.03	-0.13	-0.07	-0.02	-0.14	-0.16	0.39	-0.09	-0.04	0.11	-0.10	-0.10	-0.01	-0.20	-0.02	0.04	0.00	X			
Military Area (MA)	-0.04	-0.04	-0.03	-0.02	-0.05	-0.06	0.02	0.05	-0.06	0.01	-0.05	-0.06	-0.04	-0.04	-0.04	0.02	0.02	0.01	-0.01	-0.01	-0.04	X		
Recreational Area (RcA)	-0.02	-0.03	-0.03	-0.04	-0.02	-0.03	-0.02	0.02	-0.03	-0.03	-0.02	0.03	-0.04	-0.02	-0.02	-0.02	0.03	-0.02	-0.02	-0.01	0.00	-0.02	X	
Residential Area (RsA)	0.14	0.28	0.30	0.19	0.21	0.45	0.18	0.31	0.35	-0.08	0.41	-0.24	0.47	0.17	0.15	-0.27	-0.71	-0.12	-0.15	-0.09	-0.21	-0.11	-0.07	X

Figure 1 – Correlation matrix

Table 1 – Descriptive statistics of predicting variables

	Mean	Std. error
Touristic POI	8.22	0.946
Traffic Signals	7.52	0.496
Bus Stops	7.76	0.351
Rail/Tram Stops	2.28	0.174
Educational Points	0.72	0.081
Healthcare Points	2.07	0.171
Railway km	3.71	0.367
Road km	19.77	0.462
Capacity of BSS	3.29	0.491
Water Area	6.20	1.025
Cemetery Area	0.78	0.463
Commercial Area	3.60	0.655
Industrial Area	8.93	1.286
Military Area	0.62	0.280
Recreation Area	0.14	0.068
Residential Area	64.71	2.086
Destination Trips	89.68	15.142

3.2 Analysis methods

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;

$$y_i = \beta_{i0} + \beta_{i1} x_{i1} + \beta_{i2} x_{i2} + \cdots + \beta_{ik} x_{ik} + \varepsilon_i \tag{1}$$

where: y_i represents the observation i of the target variable; β_{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; ε_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:

$$y = \rho W y + \beta X + \varepsilon \tag{2}$$

where y is an N×1 vector of observations on the dependent variable; X is an N×K matrix of observations on the explanatory variables; β represents a K×1 vector of regression slope coefficients; ε is the error term vector, ρ 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:

$$y = \beta X + e \tag{3}$$

where $e = \lambda W e + \varepsilon$, and λ 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:

$$y_i = \beta_{i0}(u_i, v_i) + \sum_{k=1}^{n} \beta_{ik}(u_i, v_i) x_{ik} + \varepsilon_i \tag{4}$$

where: y_i is the i-th observation of the target variable; β_{ik} represents the coefficient of the k-th independent variable; x_{ik} represents the k-th independent variable; ε_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:

$$\sum_{j=1}^{n} w_{ij}(y_i - \beta_{i0} - \sum_{k=1}^{p} \beta_{ik} x_{ik})^2 \tag{5}$$

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 auto-correlation [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 2 – Goodness-of-fit for the OLS and spatial regression models

	OLS	SEM	SLM
AIC	703.38	667.25	689.17
Likelihood	-350.07	-335.82	-340.33
R-squared	0.637	0.832	0.815

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

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
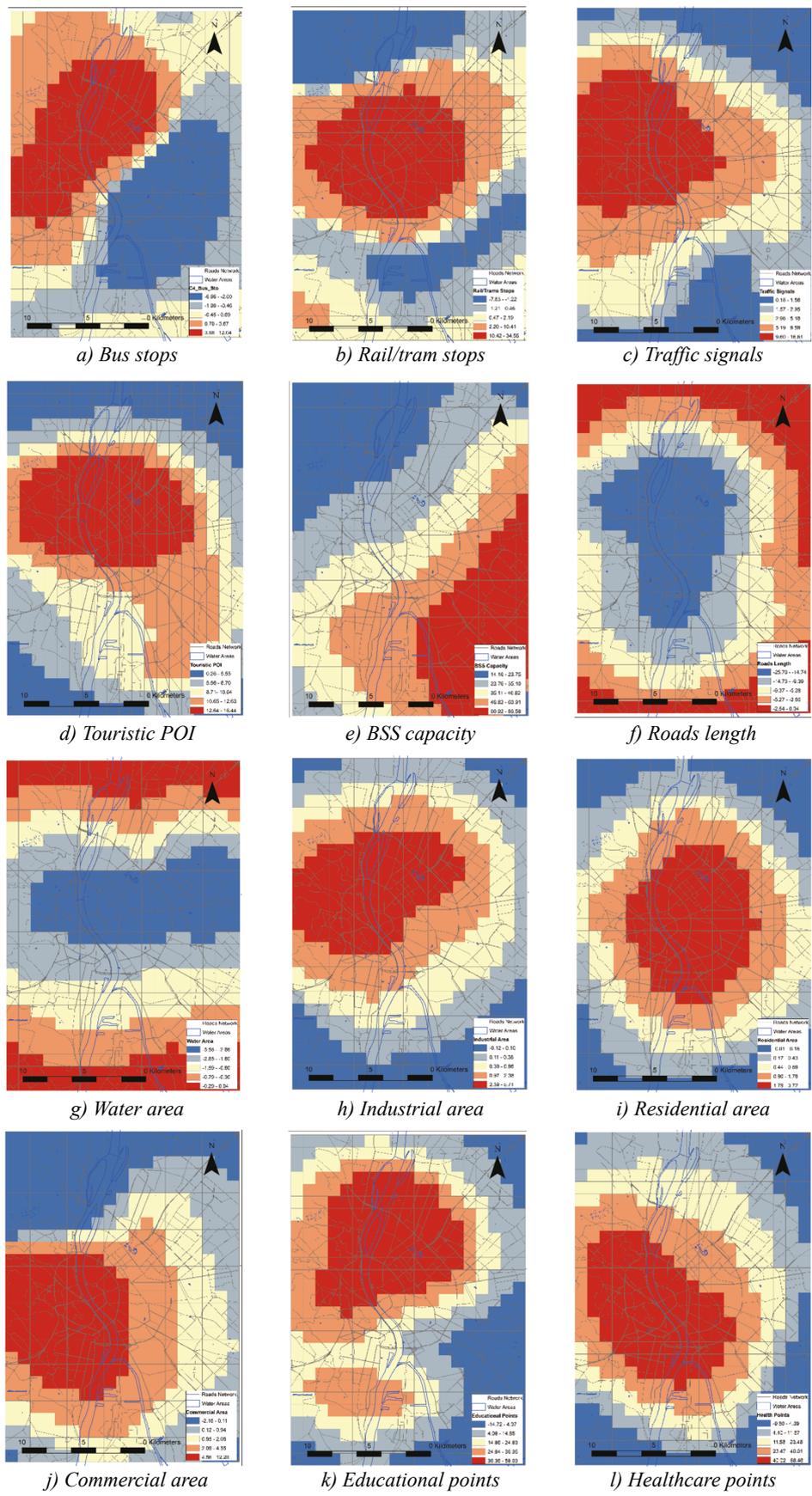
Figure 2 – Spatial distribution of the regression coefficients estimated for significant variables

Table 3 – Results estimation of the predicting variables

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

* 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 4 – Performance comparison between the training and test sets for the SEM models

Model	R-squared training set	R-squared test set	RMSE training set	RMSE test set
SEM	0.832	0.809	123.6	126.6

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 5 – Performance of selected models

Model	R-squared	RMSE
SEM	0.832	123.6
GWR	0.823	114.0
OLS	0.637	202.3
RF	0.674	229.3
Poisson	0.683	312.9
NB	0.695	312.4

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 that 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

[1] Nielsena TAS, Skov-Petersen H, Carstensen TA. Urban planning practices for bikeable cities – The case of Copenhagen. *Urban Research & Practice*. 2013;6(1):110-115. DOI: 10.1080/17535069.2013.765108.

[2] Jaber A, Juhász J, Csonka B. An analysis of factors affecting the severity of cycling crashes using binary regression model. *Sustainability.* 2021;13(12):6945. DOI: 10.3390/su13126945.

[3] Gao J, et al. Evaluating the cycling comfort on urban roads based on cyclists' perception of vibration. *Journal of Cleaner Production*. 2018;192(August 2018):531-541. DOI: 10.1016/j.jclepro.2018.04.275.

[4]   Si H, et al. Mapping the bike sharing research published from 2010 to 2018: A scientometric review. *Journal of Cleaner Production*. 2019;213(March 2019):415-427. DOI: 10.1016/j.jclepro.2018.12.157.

[5]   El-Assi W, Mahmoud MS, Habib KN. Effects of built environment and weather on bike sharing demand: A station level analysis of commercial bike sharing in Toronto. *Transportation*. 2017;44:589-613. DOI: 10.1007/s11116-015-9669-z.

[6]   Gu T, Kim I, Currie G. To be or not to be dockless: Empirical analysis of dockless bikeshare development in China. *Transportation Research Part A: Policy and Practice*. 2019;119(January 2019):122-147. DOI: 10.1016/j.tra.2018.11.007.

[7]   Pal A, Zhang Y. Free-floating bike sharing: Solving real-life large-scale static rebalancing problems. *Transportation Research Part C: Emerging Technologies*. 2017;80(July 2017):92-116. DOI: 10.1016/j.trc.2017.03.016.

[8]   Lazarus J, et al. Micromobility evolution and expansion: Understanding how docked and dockless bikesharing models complement and compete – A case study of San Francisco. *Journal of Transport Geography*. 2020;84(April 2020):102620. DOI: 10.1016/j.jtrangeo.2019.102620.

[9]   Faghih-Imani A, Elurub N. Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: A case study of New York CitiBike system. *Journal of Transport Geography*. 2016;54(June 2016):218-227. DOI: 10.1016/j.jtrangeo.2016.06.008.

[10] Giot R, Cherrier R. Predicting bikeshare system usage up to one day ahead. *2014 IEEE Symposium on Computational Intelligence in Vehicles and Transportation Systems (CIVTS)*. 2014. p. 22-29. DOI: 10.1109/CIVTS.2014.7009473.

[11] Ashqar HI, et al. Network and station-level bike-sharing system prediction: A San Francisco bay area case study. *Journal of Intelligent Transportation Systems*. 2021. DOI: 10.1080/15472450.2021.1948412.

[12] Guidon S, Reck DJ, Axhausen K. Expanding a(n) (electric) bicycle-sharing system to a new city: Prediction of demand with spatial regression and random forests. *Journal of Transport Geography*. 2020;84(April 2020):102692. DOI: 10.1016/j.jtrangeo.2020.102692.

[13] Wang X, Cheng Z, Trépanier M, Sun L. Modeling bike-sharing demand using a regression model with spatially varying coefficients. *Journal of Transport Geography*. 2021;93(May 2021):103059. DOI: 10.1016/j.jtrangeo.2021.103059.

[14] Rixey RA. Station-level forecasting of bikesharing ridership: Station network effects in three U.S. Systems. *Transportation Research Record*. 2013;2387(1):46-55. DOI: 10.3141/2387-06.

[15] Faghih-Imania A, et al. How land-use and urban form impact bicycle flows: Evidence from the bicycle-sharing system (BIXI) in Montreal. *Journal of Transport Geography*. 2014;41(December 2014):306-314. DOI: 10.1016/j.jtrangeo.2014.01.013.

[16] Mateo-Babiano I, Bean R, Corcoran J, Pojani D. How does our natural and built environment affect the use of bicycle sharing? *Transportation Research Part A: Policy and Practice*. 2016;94(December 2016):295-307. DOI: 10.1016/j.tra.2016.09.015.

[17] Noland RB, Smart MJ, Guo Z. Bikeshare trip generation in New York City. *Transportation Research Part A: Policy and Practice*. 2016;94(December 2016):164-181. DOI: doi.org/10.1016/j.tra.2016.08.030.

[18] Scott DM, Ciuro C. What factors influence bike share ridership? An investigation of Hamilton, Ontario's bike share hubs. *Travel Behaviour and Society*. 2019;16(July 2019):50-58. DOI: 10.1016/j.tbs.2019.04.003.

[19] Shen Y, Zhang X, Zhao J. Understanding the usage of dockless bike sharing in Singapore. *International Journal of Sustainable Transportation*. 2018; p. 1-15. DOI: 10.1080/15568318.2018.1429696.

[20] Bao J, Shi X, Zhang H. Spatial analysis of bikeshare ridership with smart card and POI data using geographically weighted regression method. *IEEE Access*. 2018;6:76049-76059. DOI: 10.1109/ACCESS.2018.2883462.

[21] Munira S, Sener IN. A geographically weighted regression model to examine the spatial variation of the socioeconomic and land-use factors associated with Strava bike activity in Austin, Texas. *Journal of Transport Geography*. 2020;88(October 2020):102865. DOI: 10.1016/j.jtrangeo.2020.102865.

[22] Yang H, et al. Exploring spatial variation of bike sharing trip production and attraction: A study based on Chicago's Divvy system. *Applied Geography*. 2020;115(February 2020):102130. DOI:

10.1016/j.apgeog.2019.102130.

[23] Yang F, Ding F, Qu X, Ran B. Estimating urban shared-bike trips with location-based social networking data. *Sustainability.* 2019;11(11):3220. DOI: 10.3390/su11113220.

[24] Zhao D, Ong GP, Wang W, Hu XJ. Effect of built environment on shared bicycle reallocation: A case study on Nanjing, China. *Transportation Research Part A: Policy and Practice.* 2019;128(October 2019):73-88. DOI: 10.1016/j.tra.2019.07.018.

[25] Xu Y, et al. Unravel the landscape and pulses of cycling activities from a dockless bike-sharing system. *Computers, Environment and Urban Systems.* 2019;75(May 2019):184-203. DOI: 10.1016/j.compenvurbsys.2019.02.002.

[26] Zhang Y, Thomas T, Brussel M, van Maarseveen M. Exploring the impact of built environment factors on the use of public bikes at bike stations: Case study in Zhongshan, China. *Journal of Transport Geography.* 2017;58(January 2017):59-70. DOI: 10.1016/j.jtrangeo.2016.11.014.

[27] Li B, et al. Exploring urban taxi ridership and local associated factors using GPS data and geographically weighted regression. *Cities.* 2019;87(April 2019):68-86. DOI: 10.1016/j.cities.2018.12.033.

[28] Cardozo OD, García-Palomares JC, Gutiérrez J. Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. *Applied Geography.* 2012;34(May 2012):548-558. DOI: 10.1016/j.apgeog.2012.01.005.

[29] Chiou YC, Jou RC, Yang CH. Factors affecting public transportation usage rate: Geographically weighted regression. *Transportation Research Part A: Policy and Practice.* 2015;78(August 2015):161-177. DOI: 10.1016/j.tra.2015.05.016.

[30] Jaber A, Baker LA, Csonka B. The influence of public transportation stops on bike-sharing destination trips: Spatial analysis of Budapest City. *Future Transportation.* 2022;2(3):688-697. DOI: 10.3390/futuretransp2030038.

[31] Pu Z, et al. Evaluation of spatial heterogeneity in the sensitivity of on-street parking occupancy to price change. *Transportation Research Part C: Emerging Technologies.* 2017;77(April 2017):67-79. DOI: 10.1016/j.trc.2017.01.008.

[32] Pan Y, et al. Investigating the impacts of built environment on traffic states incorporating spatial heterogeneity. *Journal of Transport Geography.* 2020;83(February 2020):102663. DOI: 10.1016/j.jtrangeo.2020.102663.

[33] Huang Y, Wang X, Patton D. Examining spatial relationships between crashes and the built environment: A geographically weighted regression approach. *Journal of Transport Geography.* 2018;69(May 2018):221-233. DOI: 10.1016/j.jtrangeo.2018.04.027.

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

[35] Obaid M, Torok A. Macroscopic traffic simulation of autonomous vehicle effects. *Vehicles.* 2021;3(2):187-196. DOI: 10.3390/vehicles3020012.

[36] Bucsky P. Modal share changes due to COVID-19: The case of Budapest. *Transportation Research Interdisciplinary Perspectives.* 2020;8(November 2020):110141. DOI: 10.1016/j.trip.2020.100141.

[37] Fraboni F, et al. A cluster analysis of cyclists in Europe: Common patterns, behaviours, and attitudes. *Transportation.* 2021. DOI: 10.1007/s11116-021-10187-3.

[38] Mátrai T, Tóth J. Cluster analysis of public bike sharing systems for categorization. *Sustainability.* 2020;12(14):5501. DOI: 10.3390/su12145501.

[39] Soltani A, Mátrai T, Camporeale R, Allan A. Exploring shared-bike travel patterns using big data: Evidence in Chicago and Budapest. In*: Computational Urban Planning and Management for Smart Cities (CUPUM 2019), Lecture Notes in Geoinformation and Cartography.* Springer, Cham; 2019. p. 53-68. DOI: 10.1007/978-3-030-19424-6_4.

[40] Koller B, Hegglin D, Schnyder M. A grid-cell based fecal sampling scheme reveals: Land-use and altitude affect prevalence rates of Angiostrongylus vasorum and other parasites of red foxes (Vulpes vulpes). *Parasitology Research.* 2019;118:2235-2245. DOI: 10.1007/s00436-019-06325-7.

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

grid-cell-based analysis using remote sensing data. *Journal of African Earth Sciences.* 2016;123(November 2016):381-391. DOI: 10.1016/j.jafrearsci.2016.08.014.

[42] Schimohr K, Scheiner J. Spatial and temporal analysis of bike-sharing use in Cologne taking into account a public transit disruption. *Journal of Transport Geography*. 2021;92(April 2021):103017. DOI: 10.1016/j.jtrangeo.2021.103017.

[43] Wu C, Kim I, Chung H. The effects of built environment spatial variation on bike-sharing usage: A case study of Suzhou, China. *Cities*. 2021;110(March 2021):103063. DOI: 10.1016/j.cities.2020.103063.

[44] Radzimski A, Dzięcielski M. Exploring the relationship between bike-sharing and public transport in Poznań, Poland. *Transportation Research Part A: Policy and Practice.* 2021;145(March 2021):189-202. DOI: 10.1016/j.tra.2021.01.003.

[45] Brunsdon C, Fotheringham AS, Charlton ME. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis*. 1996;28(4):281-298. DOI: 10.1111/j.1538-4632.1996.tb00936.x.

[46] Ma X, et al. Modeling the factors influencing the activity spaces of bikeshare around metro stations: A spatial regression model. *Sustainability.* 2018;10(11):3949. DOI: 10.3390/su10113949.

[47] Yang W, et al. The spatial characteristics and influencing factors of modal accessibility gaps: A case study for Guangzhou, China. *Journal of Transport Geography*. 2017;60(April 2017):21-32. DOI: 10.1016/j.jtrangeo.2017.02.005.

[48] Hurvich CM, Simonoff JS, Tsai CL. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2002;60(2):271-293. DOI: 10.1111/1467-9868.00125.

[49] Shen H, et al. Exploring a pricing model for urban rental houses from a geographical perspective. *Land.* 2022;11(1):4. DOI: 10.3390/land11010004.

كيف يؤثر كل من استخدام الأراضي والبيئة المبنية ومرافق النقل على وجهات رحلات مشاركة الدراجات؟

أحمد جابر*، بالينت تشونكا

يهدف هذا البحث إلى دراسة تأثير استخدام الأراضي ، والبيئة المبنية ، ومواقع مرافق النقل العام على وجهة رحلات مشاركة الدراجات في المناطق الحضرية. تمت دراسة العديد من النماذج الإحصائية لتحديد العلاقة بين المتغيرات ووجهات الرحلة ؛ مثل انحدار المربعات الصغرى العادية والانحدار المكاني والانحدار الموزون جغرافيًا. بالإضافة إلى ذلك ، تم إجراء مقارنة بين النماذج المقترحة ونماذج إضافية. تم جمع البيانات في مدينة بودابست، المجر. وضحت النتائج أن النقاط السياحية ذات الأهمية ونقاط الرعاية الصحية والنقاط التعليمية لها تأثير إيجابي على وجهات مشاركة الدراجات. تعمل محطات النقل العام في الحافلات والسكك الحديدية على جذب مستخدمي الدراجات الذين يتشاركون في استخدام الدراجات. استخدام الأراضي له تأثيرات مختلفة على وجهات رحلات مشاركة الدراجات؛ في الغالب كانتشار دائري داخل الهيكل الحضري للمدينة، مثل المناطق السكنية والصناعية والتجارية والتعليمية. تشكل المتغيرات الأخرى مثل طول الطرق ومناطق المياه كقيود على وجهات رحلة مشاركة الدراجات. يعمل الانحدار الموزون جغرافيًا والمكاني بشكل أفضل من نماذج التعداد والغابات العشوائية. تساعد هذه الدراسة صانعي القرار في التنبؤ بمصفوفة الأصل والوجهة لرحلات مشاركة الدراجات بناءً على شبكة النقل واستخدام الأراضي.

مشاركة الدراجات، النقل العام، استخدام الأراضي، التحليل المكاني، وجهات الرحلات