Accident Analysis and Prevention 132 (2019) 105259

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Macro-level accident modeling in Novi Sad: A spatial regression approach

Miloš Pljakić[a], Dragan Jovanović[b,*], Boško Matović[b], Spasoje Mićić[c]

[a] Faculty of Technical Sciences, University of Priština in Kosovska Mitrovica, Serbia
[b] Department of Transport and at the Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
[c] Ministry of Transport and Communications, Republic of Srpska, Bosnia and Herzegovina

ARTICLE INFO

Keywords:
Macro-level modeling
Traffic analysis zones
Spatial regression analysis
Transportation planning

ABSTRACT

In this study, a macroscopic analysis was conducted in order to identify the factors which have an effect on traffic accidents in traffic analysis zones. The factors that impact accidents vary according to the characteristics of the observed area, which in turn leads to a discrepancy between research and practice. The total number of accidents was observed in this paper, along with the number of motorized and non-motorized mode accidents within a three-year period in the city of Novi Sad. The models used for this analysis were spatial predictive models comprised of the classical predictive space model, spatial lag model and spatial error model. The spatial lag model showed the best performances concerning the total number of accidents and number of motorized mode accidents, whereas the spatial error model was prominent within the number of non-motorized mode accidents. The results found that increasing Daily Vehicle-Kilometers Traveled, parking spaces, 5-legged intersections and signalized intersections increased all types of accidents. The other demographic, traffic, road and environment characteristics showed that they had a different effect on the observed types of accidents. The results of this research can be benefitial to reserachers who deal with traffic engineering, space planning as well as making decisions with the aim of preparing countermeasures necessary for road safety improvement in the analysed area.

1. Introduction

In the last couple of years, accidents have been a problem on a global scale causing deaths, physical injuries or material damage in the whole world. Roughly 150 lives are lost every second and a couple of thousand people are injured in accidents around the world. A great number of accidents take place in developing and middle-income countries, where economic growth runs parallel to the increase in the degree of motorization as well as the increase in the population (WHO, 2015). The Republic of Serbia falls within the category of middle-income countries, where the problem of road safety has become apparent and ever-present in the last decade. In addition to this, the rate of road traffic deaths per 100,000 people is three times higher in Serbia compared to that of the developed countries of the European Union (Janstrup, 2017).

In order to monitor and improve road safety in the Republic of Serbia, attention has to be focused on the analysis of accidents in certain spatial entities. Monitoring the spatial distribution of accidents is very important for the efficient planning and implementation of appropriate activities so as to reduce the number of accidents. The spatial analysis of accidents can be conducted based on the two approaches which are consistent for its implementation. The first approach refers to the identification of hotspot locations, where a great number of accidents occur. The second one has to do with developing predictive models aimed at identifying the factors which affect the frequency of accidents. Both approaches can be applied from different macro, meso and micro levels in order to analyze the problems concerning road safety. Wang et al. (2016) point out that the analysis at the micro level is efficient with regard to identifying the problems of road safety at specific locations. However, the analyses of safety at the macro level enable a more efficient identification of such problems in a wider spatial area, which is useful for maintaining a long-term planning policy and improving road safety.

Macro level collision prediction models are used in a great number of studies with the aim of explaining the relations between the aggregate number of accidents and various characteristics (e.g., Wang et al., 2012; Osama et al., 2016; Guo et al., 2018; Saha et al., 2018; Wang et al., 2019). These models can be applied to assess the effects of certain measures for improving the safety of road users, finding and ranking the observed entities as well as assessing the potential number of accidents on the analyzed locations (Sawalha and Sayed, 2001). There have recently been significant improvements in developing

* Corresponding author at: Department of Transport, Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, Serbia.
E-mail address: draganj@uns.ac.rs (D. Jovanović).

https://doi.org/10.1016/j.aap.2019.105259
Received 20 April 2019; Received in revised form 10 July 2019; Accepted 31 July 2019
Available online 24 August 2019
0001-4575/

modes for predicting accidents. However, there is also a gap between the practical methods and the frontier of research (Mannering and Bhat, 2014). Contemporary models have good characteristics in terms of modeling accidents. However, one of the main limitations of these models is neglecting the characteristics of spatial data (Huang et al., 2018). Therefore, it is necessary to develop models which take into consideration the spatial autocorrelation between spatial data in order not to lead to inefficient and possibly inconsistent parameter estimates. Although numerous advanced spatial predictive models have been developed to date for different types of accidents, it is possible to identify certain spatial processes by using standard spatial models. These spatial processes can be determined by the spatial lag and spatial error model. The mentioned models determine the presence of spatial correlation depending on the variables (Álvarez et al., 2016) or the presence of spatial autocorrelation in the residuals (Anselin et al., 1996). Existing research has considered different types of accidents (e.g. Wier et al., 2009; Lee et al., 2015; Kamel et al., 2019). However, in previous studies the practical gap was observed in consideration of these spatial processes when it comes to modeling motorized and non-motorized mode accidents as well as comparing their results of modeling. Apart from this, it was established that, in the observed areas, different important factors were identified and they cannot be representative of all the areas and types of accidents. To be more precise, if the aggregate number of accidents is the same in two areas, their potential risk may vary and this depends on the patterns of the observed area.

This paper developed macro-level collision prediction models in traffic analysis zones in Novi Sad. The objectives of this paper are: (1) applying standard spatial models to identify the spatial processes in motorized and non-motorized mode accidents; (2) developing the classical spatial regression, spatial lag and spatial error model, which were examined and whose performance was evaluated; (3) investigating the variables comprising the traffic analysis zones – the level of demographic, traffic, road and environment characteristics which have an effect on the observed number of accidents. The results of this research can help are developing macroscopic models aimed to determine spatial processes in different types of accidents. In addition, the results can be used by the local authorities to plan future activities in terms of road safety.

2. Literature review

Numerous researchers have recently conducted macroscopic analyses of accidents in various spatial entities (Levine et al., 1995; Quddus, 2008; Wang et al., 2012; Abdel-Aty et al., 2013; Lee et al., 2014a; Osama et al., 2016; Cai et al., 2017b; Saha et al., 2018; Wang et al., 2019). In road safety, macroscopic spatial analysis aims to find a relationship between the frequency of accidents identified in a certain spatial unit and other features of that unit (Levine et al., 1995). Therefore, macroscopic spatial analyses do not deal with separate accidents but rather aggregate accidents in units of space (Lee et al., 2014b). Analyses at the macro level include spatial units such as countries (Aguero-Valverde and Jovanis, 2006; Huang et al., 2010b), regions (Amoros et al., 2003; Noland and Oh, 2004; Huang and Abdel-Aty, 2010a), traffic analysis districts (Abdel-Aty et al., 2013; Cai et al., 2017a), census blocks (Loukaitou-Sideris et al., 2007; Wier et al., 2009; Ukkusuri et al., 2011; Abdel-Aty et al., 2013; Wang and Kockelman, 2013) and traffic analysis zones (Ng et al., 2002; Hadayeghi et al., 2006; Abdel-Aty et al., 2011; Siddiqui and Abdel-Aty, 2012; Wang et al., 2012; Pulugurtha et al., 2013; Dong et al., 2014; Osama et al., 2016; Lee et al., 2018). Traffic analysis zones (TAZs) are spatial units which are mostly used for analyzing traffic safety at the macro level. Furthermore, in the research that Lee et al. (2014b) conducted, it was established that TAZs are the most suitable spatial units for achieving optimal results at the macro level compared to other spatial units. The basic characteristic of TAZs is that they include a wide range of spatial data which play a significant role in spatial models (Lee et al., 2014a). The previously mentioned studies have so far used the Geographical Information System (GIS) to collect and visualize spatial data which represent certain characteristics of the TAZs (Wang et al., 2012). These characteristics can vary from place to place, which opens up opportunities and potential for conducting a detailed case study.

In the existing studies, numerous authors have focused on identifying specific factors that model certain characteristics of TAZs in different spatial areas. The analysis of chosen characteristics depends on the database that is at the researcher's disposal and, apart from that, certain characteristics have clear trends within different research conducted in different regions (Abdel-Aty et al., 2013; Chen, 2015; Rhee et al., 2016; Jia et al., 2018; Wang et al., 2019). Demographic, traffic, road and environment characteristics of traffic analysis zones have often been examined.

Demographic characteristics play a significant role in conducting macroscopic analyses and the factors that are the most prominent ones are the total number of inhabitants, age groups, gender and the area of the observed spatial units. In a great number of studies, population density and size represent the exposure measurement which positively impacts the frequency of accidents (Ladron de Guevara et al., 2004; Wang et al., 2012; Yao et al., 2015; Lee et al., 2014a; Dong et al., 2014; Xu et al., 2014; Xu and Huang, 2015). The percentage of the population in relation to gender is often considered in the research. However, a group of authors have not identified a significant effect on the aggregate number of accidents in the TAZs (Rhee et al., 2016; Saha et al., 2018). On the other hand, a study conducted in the Washington area suggests that men reported a higher level of aggressive driving behavior compared to women (AAA Foundation for Traffic Safety, 2016). Consequently, there is a need to test the effect of the gender structure in other areas in order to plan the preventive activities which refer to the risky gender structure. Based on the previous studies, the need to examine the effects of different age groups has arisen, especially among the population older than 65 (Wang et al., 2012; AAA Foundation for Traffic Safety, 2016), below the age of 15 (Abdel-Aty et al., 2013; Huang et al., 2016; Rhee et al., 2016) and other age groups (Lee et al., 2014a; Saha et al., 2018). These groups were analysed in different spatial areas. The results of these studies indicated that age groups can have a different effect on traffic safety, which partially depends on the characteristics of the observed area. This is the reason why they need to be examined in other regions as well. In the other analysed areas, the effects of the dimensions of the TAZs (Chen, 2015; Wang et al., 2016), whose results proved useful for creating and analyzing a new zone system, were observed.

Previous studies considered different traffic characteristics when modeling the aggregate number of accidents. Annual average daily travel (AADT) (Wier et al., 2009; Fountas et al., 2018) and the vehicle-miles traveled (VMT) (Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Xu et al., 2015; Huang et al., 2016; Cai et al., 2017a) are exposure measures that have a positive effect on the aggregate number of accidents. Apart from this, numerous authors have examined the effects of the number of bus stops (Strauss et al., 2013; Wei and Lovegrove, 2013; Chen, 2015; Xie et al., 2019) and the number of parking spaces (Siddiqui and Abdel-Aty, 2012; Chen, 2015) on different types of accidents. However, such dependencies have not been confirmed, although it is assumed that there is greater mobility of vulnerable road users on these road facilities. By examining these characteristics, certain activities can be planned with regard to organizing public transportation and managing the parking service more efficiently. In addition, many authors have identified the factors of road infrastructure which have an effect on different types of accidents. The characteristics of infrastructure are often identified as follows: the density of road networks (Wang et al., 2012; Siddiqui and Abdel-Aty, 2012; Wang et al., 2016; Huang et al., 2016; Guo et al., 2017), the length of footways (Cai et al., 2016, 2017a), the length of cycleways (Chen, 2015; Saha et al., 2018), as well as intersections which are observed based on the number of legs (Wang et al., 2012; Rhee et al., 2016; Wang et al.,

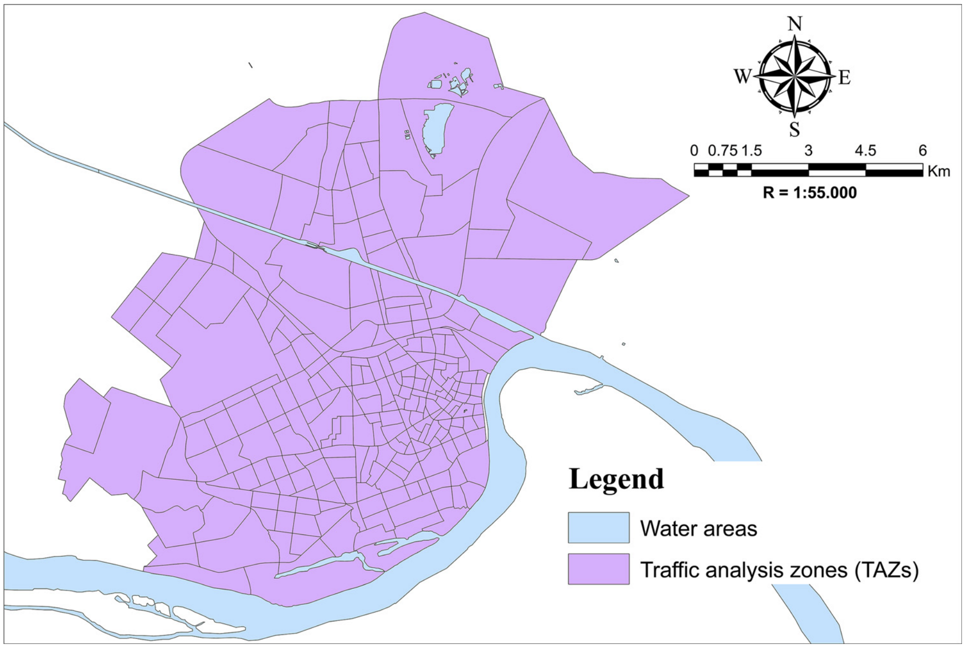

Fig. 1. The traffic analysis zones in Novi Sad.

2016) and the regulation type (Lee et al., 2015). These characteristics have a different effect on the aggregate number of accidents in relation to the observed area as well as the type of accidents, which suggests that there is a need for the given factors to be analyzed in areas where no macroscopic analyses have been conducted so far.

Until now, some authors have considered the effect of environmental characteristics in relation to the aggregate number of accidents in TAZs. Pulugurtha et al. (2013) have determined that the characteristics of an area's purpose such as urban residential, mixed use development, single-family and multi-family residential, business district and office district had a positive effect on the aggregate number of accidents. Apart from that, Levine et al. (1995) have indicated that a greater number of accidents occurred near employment centres than in residential areas. When it comes to other areas, the frequency of residential units in TAZs did not have a statistical effect on the aggregate number of accidents (Rhee et al., 2016; Saha et al., 2018). Except for the residential units, the number of schools showed a positive effect on the number of accidents (Wang et al., 2012; Rhee et al., 2016). Based on this fact, one can assess the commitment of the authorities to implementing certain activities in school zones and providing safety for children.

In the existing studies, many authors have used traditional predictive models in order to identify the effect of certain characteristics on the aggregate number of accidents. For these purposes, different predictive models were used, where the dependent variable was observed as the total number of accidents (Wang et al., 2012; Abdel-Aty et al., 2013; Cai et al., 2017b; Zhai et al., 2018; Wang et al., 2019), bicycle accidents (Siddiqui and Abdel-Aty, 2012; Lee et al., 2015; Chen, 2015; Osama et al., 2016; Saha et al., 2018), pedestrian accidents (Wier et al., 2009; Wang et al., 2016; Hong et al., 2016) as well as accidents with injury (Xu and Huang, 2015; Rhee et al., 2016; Guo et al., 2017; Xie et al., 2019). The predictive models that are most frequently used in modeling accidents are the Poisson's model (Miaou and Lum, 1993), Negative-binomial model (Poch and Mannering, 1996; Karlaftis and Tarko, 1998; Aguero-Valverde and Jovanis, 2006; Hadayeghi et al., 2006; Cai et al., 2017b) and the Poisson-lognormal model (Lord and Miranda-Moreno, 2008; Cai et al., 2017a). As far as modeling accidents at the macro level is concerned, two problems arise: (1) the spatial autocorrelation between the spatial units and (2) spatial heterogeneity in the relationships that are modeled (LeSage, 1999; Lord and Mannering, 2010). The traditional predictive models are mostly used at the micro level. However, this is not a good approach for modeling accidents at the macro level because spatial autocorrelation occurs between the spatial units. This goes against the assumption that the observed zones are independent. Dale and Fortin (2002) point out that spatial autocorrelation occurs when the observations in neighboring fields have similar data values. The contemporary approach takes spatial autocorrelation into account. Spatial non-stationarity in the literature is available for multivariate settings (Wang and Kockelman, 2013) as well as developing the conditional autoregressive model (Quddus, 2008). However, Zhou et al. (2016) pointed out that the application of the Bayesian estimation methods and the criterion for convergence is subjective, which is one of the limitations of advanced methods. Bearing in mind the mentioned limitations, it is necessary to develop regression models which use a spatial weight matrix and maximum likelihood estimation. So far, different models have been developed, but the classical spatial regression models (OLS) have been understood to be practical as they can be of use to agencies for traffic planning and safety (Rhee et al., 2016). Additionally, in areas in which a macro analysis was not conducted, there is still a practical gap, so classical spatial models are satisfactory for the first steps of improving road safety.

3. Data preparation

The data used in this study were collected in the area of Novi Sad, the capital of the Autonomous Province of Vojvodina. According to the census from 2011, there are 277,522 inhabitants in this city, which makes it the second largest in the Republic of Serbia. The administrative area of this city is 702 km^2 of which 129 km^2 is urban. Within this research, 248 TAZs geopositioned on 132 km^2 were included (Fig. 1). The spatial data regarding TAZs were taken from administrative services which deal with transportation planning in the area of Novi Sad. The division into the traffic zones was carried out by the administrative services for transport planning according to the previously defined criteria (Baass, 1981): homogeneous characteristics for each zone; recognizing natural, political, physical and historical boundaries; devising a zonal system in which the population and number of households are nearly equal in each zone; minimizing the number of intra-zonal trips; basing zonal boundaries on census zones;

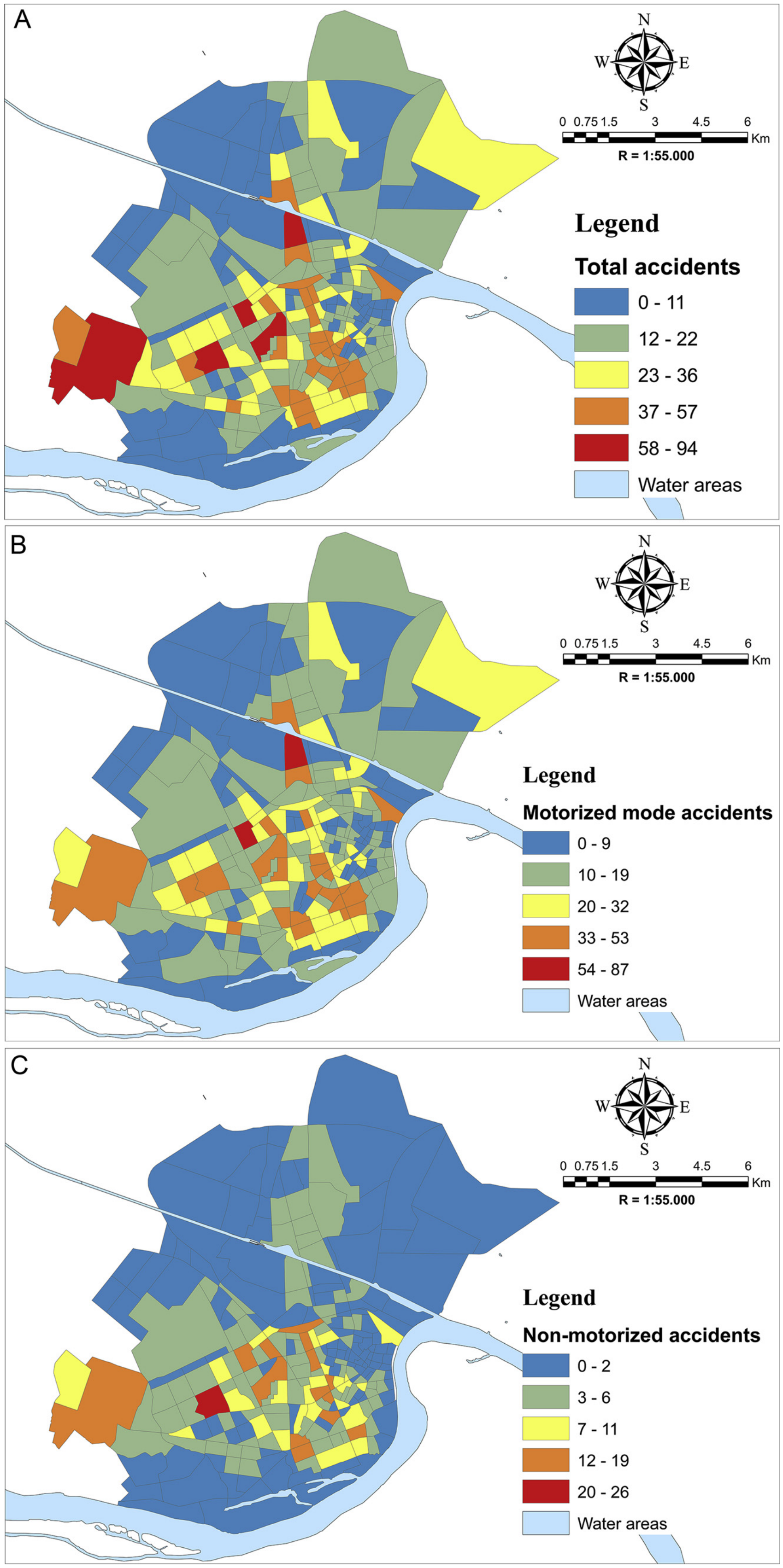

Fig. 2. Spatial distribution observed accidents in Novi Sad (2015–2017), (a) total number of accidents, b) motorized mode accidents, c) non-motorized mode accidents.

generating only connected zones; avoiding zones that are completely contained within another zone; basing zonal boundaries on census zones.

This study observes accidents in a three-year period (2015–2017) which included the total number of accidents as well as motorized and non-motorized mode accidents (Fig. 2). The data regarding the accidents were taken from a database which is authorized by the Road Traffic Safety Agency (RTSA) organized in accordance with the Common Accident Data Set (CADaS) protocol developed by the European Commission. The RTSA database is based on GIS, so that all accidents have coordinates (x,y) of the locations that are precise enough to establish the direction of traveling on major roads. On the other hand, the boundary of the TAZ units lies in the middle of the median road island or on the median line on the roads, which facilitates assessing the boundary effects for the aggregate number of accidents. An identical procedure of aggregating the number of accidents was also applied in Seoul, where the accidents in one direction belong to one TAZ and the accidents in the other direction belong to the other TAZ (Rhee et al., 2016).

The independent variables in this research are the features which describe demographic, traffic, road and environment characteristics. The demographic characteristics that were observed are the following: the area of the TAZs, the population ratio younger than 15 years old, the population ratio older than 65, the population ratio of males, the population ratio of females and population density. The traffic characteristics were represented through the following variables: the Daily Vehicle-Kilometers Traveled (DVKT), the number of bus stops, the number of parking spaces with payment (a time limit of 2 h) and the number of parking spaces with payment and no time limit. Road characteristics refer to the characteristics of urban road networks comprised of the following variables: the length of street networks, the length of footways, the length of cycleways, the number of 3-legged, 4-legged and 5-legged intersections and the number of signalized intersections. The variables used in this research were collected based on the available set of data provided by the Transport Administrative Services. Table 1 shows the descriptive statistics of the variables included in the analysis.

4. Methodology

In this research, classical spatial autoregressive models were applied in order to identify the factors that have an effect on the total number of accidents along with motorized and non-motorized mode accidents. Accidents within a specific spatial entity are discrete and represent non-negative integer values, so that countable data models are more suitable. These models are particularly useful when small spatial units are analyzed, such as segments, intersections, tunnels and road curvatures. In other words, they are suitable in locations where the frequency of accidents is very low. However, existing research has emphasized that with the analysis of spatial units that aggregate tens of accidents, the practical use of countable models becomes negligible (Rhee et al., 2016). So far, many advanced types of spatial predictive models have been developed. However, Venkataraman et al. (2014) emphasised that planning agencies still apply them inadequately. Based on this fact, classical spatial autoregressive models, the spatial lag model and spatial error model were applied in order to analyze the new set of data in the paper at hand.

4.1. Spatial autocorrelation

Spatial autocorrelation analyses the degree of connection between certain features of spatial units and features on neighboring spatial units. In this case, it is possible to establish whether the features of the traffic analysis zone affect the value of the other features in the neighboring zone (Black and Thomas, 1998). If there is spatial autocorrelation, the basic assumptions of classical regression models are violated and the data should not be analyzed using normal regression analysis. In that case, the spatial weight matrix and maximum likelihood estimation are taken into consideration in order to minimize that bias. The spatial weight matrix contains all the non-negative matrices $W = (w_{ij} : i, j = 1, \ldots, n)$ that represent the relationship between the observed spatial units (Anselin, 2001). The members of the spatial

Table 1
Descriptive statistics of the variables.

Variables	Mean	Std. Deviation	Min	Max
Accidents				
Total number of accidents	20.121	15.711	0	94
Number of motorized mode accidents	16.387	12.542	0	87
Number of non-motorized mode accidents	3.734	4.216	0	26
Demographic characteristics				
Area (ares)	5334.8	9921.4	255.9	75523.9
Population ratio younger than 15 years old (%)	12.565	5.905	0	44.44
Population ratio older than 65 years old (%)	15.137	7.547	0	42.86
Population ratio of males (%)	44.339	15.563	0	100
Population ratio of females (%)	47.597	16.388	0	100
Population density (inhabitants per m2)	0.005	0.005	0	0.020
Traffic characteristics				
DVKT	2175.794	1595.176	43.870	8071.235
Number of bus stops	0.839	1.083	0	6
Number of parking spaces with payment and a time limit of 2 h	3.887	14.187	0	96
Number of parking spaces with payment and no time limit	21.060	56.151	0	383
Road characteristics				
Length of street networks (m)	4356.4	5187.4	134.8	46553.4
Length of footways (m)	1509.3	3566.1	0	49236.7
Length of cycleways (m)	268.2	471.2	0	2724.03
Number of 3-legged intersections	5.056	4.885	0	49
Number of 4-legged intersections	3.298	2.726	0	19
Number of 5-legged intersections	0.032	0.199	0	2
Number of signalized intersections	1.137	1.179	0	5
Environment characteristics				
Number of residential units	548.149	674.231	0	4774
Density facilities (facilities per are)	0.031	0.024	0	0.117
Number of schools	0.371	0.685	0	3
Urban zones (1 - YES)	0.859	0.349	0	1

weight matrix w_{ij} represent the spatial effect of unit j onto unit i.

$$W = \begin{bmatrix} 0 & w_{ij} & w_{ik} \\ w_{ji} & 0 & w_{jk} \\ w_{ki} & w_{kj} & 0 \end{bmatrix};\ w_{ij} = \frac{1}{w_{max}} > 0 \tag{1}$$

Matrices that are based on sharing a common border play an important role in examining spatial autocorrelation. The indicator of this matrix is reflected through its members which show spatial integration and whether the TAZs share a border (l_{ij}) with another TAZ. In this paper, the rook method devised by Anselin and Grifth (1988) was applied for creating the necessary spatial weight matrices. Based on the rook method, the spatial weight matrices were defined as follows:

$$\boldsymbol{w}_{\mathbf{ij}} = \begin{Bmatrix} \mathbf{1}, & \mathbf{l_{ij} > 0} \\ \mathbf{0}, & \mathbf{l_{ij} = 0} \end{Bmatrix} \tag{2}$$

The presence of spatial autocorrelation is expressed through Moran's I index, which is one of the basic measures of spatial autocorrelation with regard to spatial data (Moran, 1948). This index contains the spatial weight matrix at its core and can be expressed through the following equation:

$$I = \frac{n}{\sum\sum_{ij} W_{ij}} \bullet \frac{\sum\sum_{ij} W_{ij}(x_i - \bar{x})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2} \tag{3}$$

Where:

x_i – is the value of the variable x on the observed unit i,
$\bar{x}$ – is the mean value of the variable x,
n – is the number of observed units,
W_{ij} – is the spatial weight matrix of distance.

The value of Moran's I index ranges from -1 to 1, where the positive/negative value suggests the positive/negative spatial autocorrelation or the appearance of clusters in an observed area. In addition to this, the Z-value can be used to assess Moran's I index, where the value is higher than 1.96 or lower than −1.96, which indicates that there is spatial autocorrelation with a confidence level of 95%.

4.2. Spatial regression models

This section of the paper deals with models through which the spatial dependence of the number of accidents and certain characteristics of traffic analysis zones can be modeled. These models take spatial autocorrelation into account in tests of statistical significance. Spatial dependence in autoregressive models is directly modeled in the regression equation (Anselin, 2001; Anselin et al., 2006). The paper observes TAZs as spatial entities for modeling accidents. The spatial effect between two zones is two-way and two-dimensional since the values of the neighboring zones have an effect on the observed zones. Based on this, spatial dependence was expressed in the following function (LeSage, 1999):

$$y_i = f(y_j),\ \text{where}\, i, j = 1, \ldots, n.;\, i \neq j, \tag{4}$$

where y_i refers to the number of accidents per TAZ. Anselin and Griffith, 1988 pointed out spatial heterogeneity, which refers to the variation in space through a classical regressive model:

$$\boldsymbol{y_i = X_j\beta_j + \varepsilon_i,\ wherei, j = 1, \ldots, n;\ i \neq j,} \tag{5}$$

where y_i is the dependent variable, X_j is the collection of explanatory variables, β_j represents the regression coefficients and ε_i is the random error whose mean value is $E[\varepsilon_i] = 0$, and the variance is constant $Var[\varepsilon_i] = \sigma^2$. Based on the classical predictive model (5), spatial autoregressive (SAR) models were introduced to model cross-sectional spatial data, which can be presented in matrix form:

$$\varepsilon \sim N(0, \sigma^2 I_n) \tag{6}$$

where y is the vector of the dependent variable (n x 1), X is the matrix of independent variables (n x k) which contains k variables. W_1 and W_2 are spatial weight matrices (n x n), ρ i λ are spatial autoregressive coefficients, β is the vector of regressive coefficients (n x 1), ε is the random error whose mean value is $E[\varepsilon_i] = 0$, and the variance is constant $Var[\varepsilon_i] = \sigma^2$ in the identity matrix I_n.

The spatial lag model (SLM) enables the presence of spatial lagged variable Cy in the model. In this model, the endogenous problem occurs since the spatially lagged value of y is in correlation with the stochastic disturbance. To be more precise, in this model the dependent variable in the observed area is subject to spill-over effects with dependent variables in neighboring areas. This can be presented in the following way:

$$\varepsilon \sim N(0, \sigma^2 I_n) \tag{7}$$

where y, X, β, ε have the same meaning as in the previous model (6). Cy is the spatially lagged dependent variable for the weights matrix.

The spatial error model (SEM) was shown through the maximum likelihood method where the disturbances display spatial dependence:

$$\varepsilon \sim N(0, \sigma^2 I_n) \tag{8}$$

where y, X, β, ε have the same meaning as in the previous model (6). W is the spatial weight matrix and λ is the coefficient on the spatially correlated errors, which has similarities to the serial correlation problem in the time series models. LeSage (1999) points out that if Moran's I index identified spatial autocorrelation, the SEM model would be suitable for the continuation of modeling.

A comparison of the applied models was conducted by the Akaike Information Criterion (AIC), Bayesian information criterion (BIC) and log-likelihood ratio. AIC represents the relationship between the log-likelihood function and number of unknown parameters i.e. the number of parameters which need to be assessed. The lowest value of the AIC and BIC is the best result of the model. The log-likelihood ratio test is a statistical test for comparing the goodness of fit of two or more statistical models (Vuong, 1989).

5. Results

Based on the specified methodology and available data, the types of spatial predictive models that model the total number of accidents and the number of motorized and non-motorized mode accidents in the area of Novi Sad were calibrated in this study. The demographic, traffic, road and environment characteristics were observed in this paper as independent variables. Accidents were a dependent variable in a model on which autocorrelation between the observed TAZ units was tested.

As far as modeling accidents is concerned, a large number of existing studies have confirmed spatial autocorrelation and indicated that the models that take account of spatial autocorrelation have better performance than the models that do not consider spatial dependence (Aguero-Valverde and Jovanis, 2006; Siddiqui et al., 2012; Wang et al., 2012; Zeng and Huang, 2014; Wang et al., 2016; Cai et al., 2016; 2017). In accordance with that, the coefficient of spatial autocorrelation was examined in this study using Moran's test. The absolute value of Moran's index ranges between 0 and 1, where the greater value suggests a degree of relationship between the observed entities and, on the other hand, the entities are observed as random spatial patterns. Fig. 2 shows the results of Moran's test for dependent variables in this research.

As presented in Fig. 3, statistically significant spatial autocorrelation (p-value < 0.05) was identified for all dependent variables. The coefficient of spatial autocorrelation was different for all variables. The lowest value was shown for the number of motorized mode accidents (Moran's I = 0.27754) and the total number of accidents (0.319954) while the highest one was for the number of non-motorized mode accidents (Moran's I = 0.342642). These results confirmed the initial hypothesis that accidents are autocorrelated in TAZs, so it is necessary to include spatial autocorrelation in the process of modeling accidents.

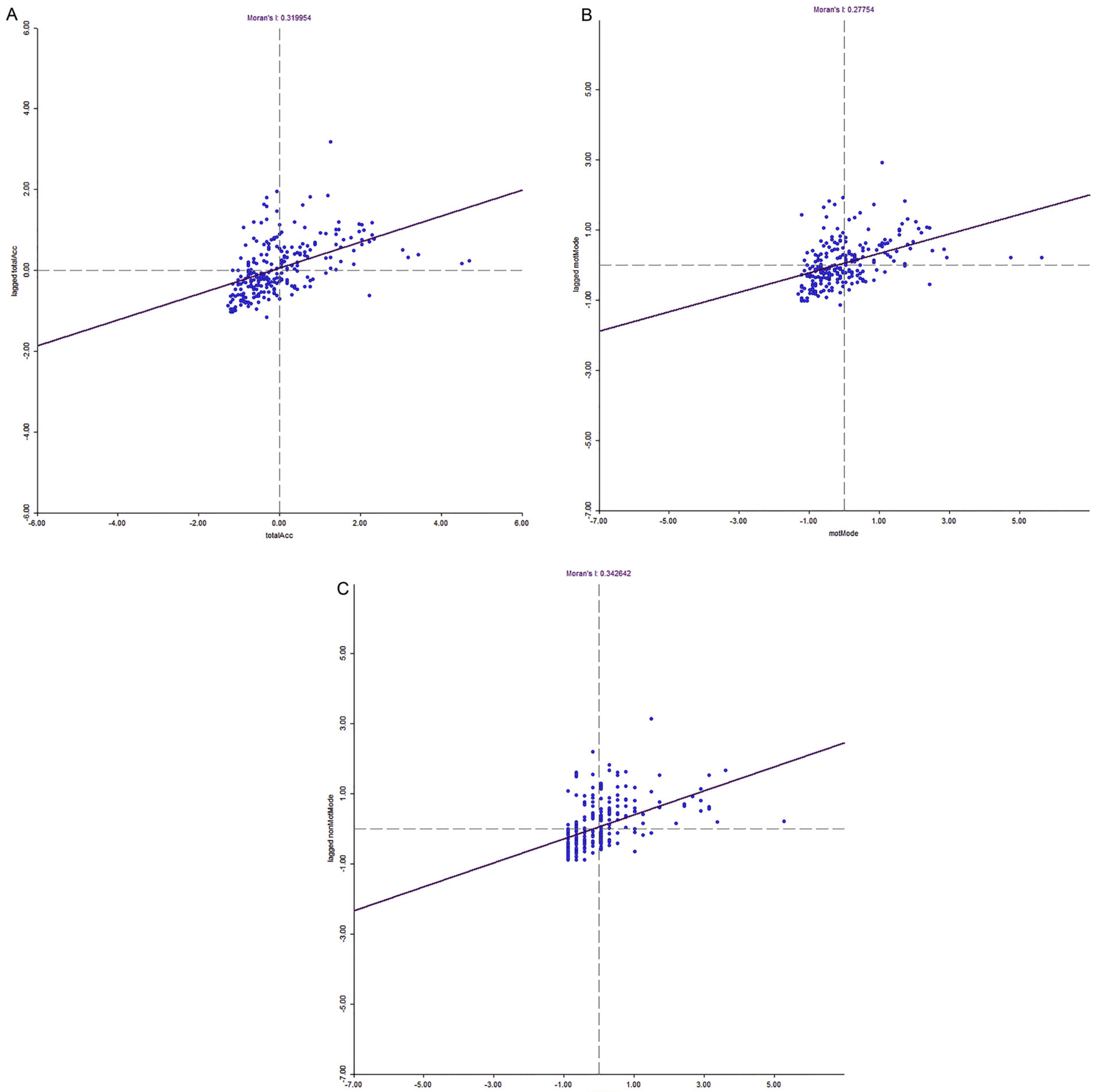

Fig. 3. Scatter plots with Moran's coefficients for dependent variables (a) total number of accidents, b) motorized mode accidents, c) non-motorized mode accidents.

Before developing the model, highly correlated variables were identified. They were included so as to avoid biased results. The Pearson correlation coefficient was measured for every pair of variables. Subsequently, the variables which showed strong interdependence and multicollinearity were excluded. This process leads to a decrease in the number of explanatory variables in advanced models (MacNab, 2004). The variables that were excluded from the process of developing the models are the following: the percent of the population younger than 15 years old, the length of footways, the number of 3-legged intersections, the density of facilities and urban zones. Finally, in the process of developing the model, all independent variables which do not have a direct or indirect effect on the aggregate number of accidents were excluded.

5.1. Total accidents models

Table 2 shows the regressive models (OLS, SLM and SEM) which indicate that there are statistically significant factors that had an effect on the total number of accidents in the observed area. Based on the Akaike information criterion (AIC), Log likelihood and R-squared model, the SLM model had the best predictive performance for modeling the total number of accidents compared to the other two models. According to the SLM model, a total of 9 significant variables were identified. In the model, DVKT shows of the exposure measure with a positive effect on the total number of accidents ($\beta = 0.004$; $p < 0.01$), which was expected. In addition, the following variables were positively impacted: the number of parking spaces with payment and a time limit of 2 h ($\beta = 0.139$; $p < 0.05$), the population of males ($\beta = 0.159$; $p < 0.01$), the length of street networks ($\beta = 0.0006$; $p < 0.01$), 5-legged intersections ($\beta = 4{,}778$; $p < 0.1$), signalized intersections ($\beta = 2.71$; $p < 0.01$) and residential units per zone ($\beta = 0.005$; $p < 0.01$). A negative effect on the total number of accidents was identified in the area traffic zones ($\beta = -0.000003$; $p < 0.01$) and the number of schools per traffic zone ($\beta = -3.213$; $p < 0.01$). Compared to the OLS model, the factor pertaining to the length of cycleways did not prove to have a statistically significant effect on the dependent variable. In addition, the model excluded the variables which do not have a direct or indirect effect on the dependent variable. Those

Table 2
Total accident model results.

	OLS		SLM		SEM	
	Coeff	SE	Coeff	SE	Coeff	SE
Demographic characteristics						
Area (are)	−0.000003***	0.000001	−0.000003***	0.000001	−0.000003**	0.000001
Population ratio of males	0.162***	0.042	0.159***	0.041	0.141***	0.040
Traffic characteristics						
Daily vehicle-kilometers traveled	0.004***	0.001	0.004***	0.001	0.004***	0.001
Number of parking spaces with payment and a time limit of 2 h	0.141**	0.046	0.139**	0.044	0.131**	0.049
Road characteristics						
Length of street networks	0.001**	0.0002	0.0006***	0.0002	0.0006**	0.0002
Length of cycleways	0.003*	0.002				
Number of 5-legged intersections			4.778*	3.128		
Number of signalized intersections	2.901***	0.709	2.71***	0.685	2.959***	0.725
Environment characteristics						
Number of residential units	0.006***	0.002	0.005***	0.0011	0.056***	0.0012
Number of all schools	−3.2928**	1.001	−3.213***	0.959	−2.878**	0.949
Constant	−3.625	2.276	−6.267	2.366	−2.654	3.107
Spatial lag coefficient			0.179	0.062		
Spatial error coefficient					0.263	0.085
No of parameters						
R-squared	0.61		0.63		0.62	
AIC	1856.84		1850.58		1854.28	
BIC	1895.49		1892.74		1892.93	
Log likelihood	−917.42		−913.29		−916.14	

Note: Variables are significant at 99% (***), 95% (**) and 90% (*) confidence level.

variables are the following: the population ratio older than 65 years old, the population ratio of females, population density, the number of bus stops, the number of parking spaces with payment and no time limit and the number of 4-legged intersections. Based on the significant variables presented in the model, it explained 63% (R-squared = 0.63) of the variability of the responses around its mean.

5.2. Models of motorized mode accidents

The results of the statistically significant variables in models of motorized mode accidents are shown in Table 3. The results were similar to those models of the total number of accidents. According to the Log likelihood and AIC criterion, the SLM model also proved to have the best performances in modeling motorized mode accidents. In the SLM model, 10 statistically significant variables which have an effect on motorized mode accidents were identified. With these types of accidents, the exposure measure is the DVTK variable with an expectedly positive effect, as with the total number of accidents ($\beta = 0.003$; $p < 0.01$). Other variables with a positive effect were similar to the ones in the models with the total number of accidents, except for 5-legged intersections per zones, which were not statistically significant in the model. Regarding the variables with a negative effect on accidents, the area of traffic zones ($\beta = -0.000003$; $p < 0.01$) and the number of schools per traffic zone ($\beta = -2.782$; $p < 0.01$) were also identified as with the total number of accidents. Additionally, the population ratio of females ($\beta = -0.075$; $p < 0.05$) and older than 65 ($\beta = -0.008$; $p < 0.1$) also showed a negative effect, which is typical of this age group in the observed area. The variables that did not have a

Table 3
Motorized mode accidents models results.

	OLS		SLM		SEM	
	Coeff	SE	Coeff	SE	Coeff	SE
Demographic characteristics						
Area	−0.000003**	0.000001	−0.000003***	0.000001	−0.000003**	0.000001
Population ratio older than 65			−0.008*	0.005	−0.0085*	0.005
Population ratio of males	0.1668***	0.039	0.167***	0.038	0.156***	0.038
Population ratio of females	−0.070*	0.005	−0.075**	0.047	−0.076**	0.048
Traffic characteristics						
Daily vehicle-kilometers traveled	0.003***	0.001	0.003***	0.001	0.003***	0.0001
Number of parking spaces with payment and a time limit of 2 h	0.116**	0.039	0.115**	0.037	0.108**	0.041
Road characteristics						
Length of street networks	0.0005**	0.0002	0.0005***	0.0002	0.0005**	0.0001
Number of signalized intersections	2.604***	0.606	2.443***	0.587	2.635***	0.621
Environment characteristics						
Number of residential units	0.0045***	0.001	0.004***	0.00098	0.0043***	0.0001
Total number of schools	−2.841***	0.863	−2.782***	0.829	−2.558***	0.819
Constant	−3.298	2.265	−2.424	2.316	−1.957	2.609
Spatial lag coefficient			0.176	0.065		
Spatial error coefficient					0.251	0.086
R-squared	0.55		0.57		0.56	
AIC	1778.66		1772.70		1776.32	
BIC	1817.31		1814.86		1814.97	
Log likelihood	−878.33		−874.35		−877.16	

Note: Variables are significant at 99% (***), 95% (**) and 90% (*) confidence level.

Table 4
Non-motorized mode accidents models results.

	OLS		SLM		SEM	
	Coeff	SE	Coeff	SE	Coeff	SE
Demographic characteristics						
Area	−0.000001*	0.000000				
Population density			−95.43*	57.676	−119.98**	62.10
Traffic characteristics						
Daily vehicle-kilometers traveled	0.0008***	0.0002	0.0007***	0.0001	0.0007***	0.0002
Number of bus stops	0.353*	0.194	0.317*	0.185	0.313*	0.186
Number of parking spaces with payment and a time limit of 2 h	0.0345**	0.013	0.032**	0.013	0.031**	0.014
Road characteristics						
Length of cycleways	0.0011**	0.0005	0.0010**	0.0004	0.0012**	0.0005
Number of 5-legged intersections	2.272**	0.989	2.249***	0.943	1.852**	0.968
Number of signalized intersections	0.481**	0.214	0.448**	0.204	0.555**	0.218
Environment characteristics						
Number of residential units	0.0023***	0.0005	0.0022***	0.0004	0.003***	0.0005
Constant	−1.829	0.935	−0.821	0.413	−1.066	0.929
Spatial lag coefficient			0.219	0.068		
Spatial error coefficient					0.345	0.081
R-squared	0.52		0.55		0.57	
AIC	1252.85		1243.96		1238.8	
BIC	1287.97		1282.61		1273.91	
Log likelihood	−616.42		−610.98		−609.39	

Note: Variables are significant at 99% (***), 95% (**) and 90% (*) confidence level.

direct or indirect effect on the dependent variable are the following: population density, the number of bus stops, the number of parking spaces with payment and no time limit, the length of cycleways, the number of 4-legged intersections and the number of 5-legged intersections. The value of the R-squared indicated that the presented model explained 57% (R-squared = 0.57) of the variability of the responses around its mean.

5.3. Models of non-motorized mode accidents

Non-motorized mode accidents included accidents in which vulnerable road users in the observed area were involved. The results of these types of accidents are presented in Table 4. In comparison with the models for other types of accidents, the SEM model had the best performances in modeling non-motorized mode accidents. In the SEM model, 7 statistically significant variables which had an effect on non-motorized accidents were identified. Among the more significant variables, a positive effect of the DVKT variable ($\beta = 0.0007$; $p < 0.01$) was identified. This variable refers to the exposure measure, as with the previous models. Unlike the total number of accidents, a positive effect on the non-motorized mode accidents was identified with the variables that refer to the number of bus stops ($\beta = 0.313$; $p < 0.1$) and length of cycleways ($\beta = 0.0012$; $p < 0.05$) that greatly characterize these types of accidents. In addition, when it comes to non-motorized mode accidents, a negative effect of population density ($\beta = -119.98$; $p < 0.05$) was present, which indicated the specificity of the observed area. Those variables that did not show a statistically significant effect on the dependent variable were excluded from the model. The excluded variables are the following: the population ratio older than 65 years old, the population ratio of males, the population ratio of females, the length of street networks, the number of 4-legged intersections, the number of parking spaces with payment and no time limit and the number of schools. With regard to the R-squared values, the SEM model explained 57% of the variability of the responses around its mean.

6. Discussion

A new set of available data were analyzed in this research in order to assess the effect of demographic, traffic, road and environment characteristics on the frequency of accidents within the TAZs. The new set of data included a set of independent factors and three dependent variables which made up for the total number of accidents and the number of motorized and non-motorized mode accidents. The effect of the independent factors was examined by developing three models (i.e. OSL, SLM and SEM). The variables which were significant in the model that showed the best performances for different types of accidents were discussed. Taking into account the total number of accidents and motorized mode accidents, the SLM model displayed the best performances, whereas the SEM model was significant for non-motorized mode accidents. With all types of accidents in the observed area, spatial autocorrelation was established between the traffic zones, which was expected and in conducted with the existing studies conducted on the territory of the following cities: Hillspore and Pinellas (Siddiqui et al., 2012; Abdel-Aty et al., 2013), Shanghai (Wang et al., 2016, 2019), Hong Kong (Guo et al., 2017), Seoul (Rhee et al., 2016), Seattle (Chen, 2015) and Suzhou (Jia et al., 2018).

The demographic characteristics included six independent variables: population density, the population ratio of males, females, younger than 15, older than 65 and area traffic zones. In the observed area, population density had a negative influence on non-motorized mode accidents. This meant that in zones with great population density, there were a smaller number of accidents with vulnerable road users. In addition to this, great population density in zones led to an increase in the population mobility as a results of which the effect of Safety-In-Numbers occurred, which proved to exist in many urban areas (Leden, 2002; Jacobsen, 2015; Elvik et al., 2013; Elvik and Bjørnskau, 2017; Tasic et al., 2017; Xie et al., 2018; Lee et al., 2019). Apart from population density, the population ratio was considered in accordance with its gender and age groups. The results revealed that in those zones with a higher population ratio of males, the total number of accidents as well as the number of motorized mode accidents increased. This can be explained by the fact that males are participants in traffic as drivers of motorized vehicles more than women (Santamariña-Rubio et al., 2014). Furthermore, the effect of the male gender can be explained by the fact that the population ratio of females in the zones had a negative effect on the number of motorized mode accidents. As far as the age groups are concerned, the population ratio younger than 15 did not have an effect on the examined dependent variables. This finding is in line with the previous study conducted by Hillspore and Pinellas (Abdel-Aty et al., 2013). However, the population ratio older than 65 showed a negative effect on the number of motorized mode accidents in the zones. These results can be explained by the fact that people older than

65 take part less frequently in accidents characterized by losing control or exceeding the speed limit, risky overtaking or driving under the influence of alcohol (Hakamies-Blomqvist et al., 1993). In addition to this, the older population is more aware of the risk that has to do with exceeding the speed limit and driving under the influence of alcohol and that they are more willing to avoid such risky behavior, which often leads to motorized mode accidents. Also, older people are less frequently drivers in traffic because of their psycho-physical limitations (Koppel et al., 2019). Similar results were recorded in the area of Central Florida (Lee et al., 2015). The final trait of demographic characteristics observed in this paper is the area of the TAZ which recorded a negative effect on the total number of accidents and number of motorized mode accidents. Based on the results, it can be concluded that the area of the TAZ was linked to population density so that in the central parts of the city there were smaller areas of TAZs. However, there was an increase in the total number of accidents and number of motorized mode accidents. This was not the case when the number of non-motorized mode accidents was observed, where the area did not have a statistically significant effect. Based on the analyzed demographic characteristics, it is necessary to apply the adequate programs and activities which are aimed at educating the population, especially males in the observed area.

The traffic characteristics included were the DVKT, the number of bus stops, the number of parking spaces with payment and a time limit of 2 h as well as the number of parking spaces with payment and no time limit. The DVKT variable was the measure of exposure which had a positive effect on all three groups of accidents included in the study. In all the models, this variable was identified as statistically significant, which suggests that if the traffic flow on road networks increased within one zone, there was an increase in the number of accidents. The existing studies have observed the vehicle-miles traveled, which also had a positive effect on the accidents (Abdel-Aty et al., 2013; Wang and Kockelman, 2013; Xie et al., 2019). Additionally, the number of bus stops showed a positive effect on the number of non-motorized mode accidents. The effect of this factor can be explained by the fact that the presence of bus stops led to higher population mobility, especially of pedestrians, where accidents involving pedestrians often occurred. Apart from bus stops, this paper analyzed the number of parking spaces for vehicles which were charged within two zones. The zones were divided into the red zone, which included the number of parking spaces with payment (time limit of 2 h), and the blue zone, which included the number of parking spaces with payment and no time limit. The number of parking spaces with payment and no time limit was not statistically significant for the observed accidents. However, the number of parking spaces with payment and a time limit of 2 h did not show statistical significance with all the dependent variables. The number of parking spaces with payment and a time limit of 2 h had a positive effect on the number of all accidents and the number of motorized and non-motorized mode accidents. This finding can be explained in two different ways. The first one is that on those parking spaces with a limited usage time, there was a frequent change in the vehicles, which led to unexpected situations both for the drivers of motorized vehicles and vulnerable road users. The second approach had to do with the fact that parking spaces with a time limit were often distributed in central parts of the city with popular landmarks, which often led to complex traffic situations. These findings are very useful for the traffic services on the territory of Novi Sad in terms of planning and creating more efficient infrastructure facilities intended for the public transportation of passengers and parking service.

The road characteristics included seven variables that referred to the length of the street networks, facilities for the transportation of pedestrians and cyclists and the number of intersections. The total length of roads in an observed area had a positive effect on the total number of accidents as well as the number of motorized mode accidents. This variable can be observed as a measure of exposure since it was directly associated with the frequency of accidents. These findings were consistent with the previous research in which the effect of the length of roads on the accidents within the TAZs was identified (Cai et al., 2017a). Apart from the total length of the roads, the length of the cycling paths also had a positive effect on the number of non-motorized mode accidents, which was consistent with other research (Lusk et al., 2011; Wei and Lovegrove, 2013). This can be explained by the fact that the cycleways next to the roads increased the number of potential conflicts, which in turn led to an increase in the likelihood of bicycle accidents. Additionally, intersections showed a positive effect on the number of non-motorized mode accidents i.e. the number of signalized intersections and number of 5-legged intersections. The number of signalized intersections had a statistically significant effect on all dependent variables which confirmed the fact that there were late decisions of drivers because of dilemma zones. This led to a great number of accidents (Wu, 2014). In addition to this, the number of 5-legged intersections hindered the movement of vulnerable road users, especially when crossing the road, which required much greater efforts than with 3-legged intersections and 4-legged intersections. However, in this study, the number of 3-legged and 4-legged intersections did not have a statistically significant effect on accidents that occurred in the observed area.

Within the environment characteristics, this paper deals with the residential units, the density of all facilities, schools and a dummy variable of urban zones, which reflects being in an urban environment (1 – yes; 0 – no). The number of residential units in TAZs had a positive effect on all the dependent variables observed in this study. This can be explained by the fact that in the observed areas, the density of residential units was higher in central parts than in the periphery, which is also common for the frequency of traffic accidents. In central parts, there was greater population mobility, which suggests that all categories of the participants in traffic were exposed to accidents. In certain areas, the effect of the residential units on the total number of accidents was not confirmed (Abdel-Aty et al., 2013; Rhee et al., 2016). However, when it comes to vulnerable road users, certain studies have confirmed the effect of residential units on the frequency of accidents with pedestrians (Siddiqui et al., 2012) and cyclists (Siddiqui et al., 2012; Chen, 2015). Furthermore, the density of all facilities and the locations of the TAZs in relation to the gravitation area of the city were not statistically significant, which is in line with the previous studies (Siddiqui et al., 2012; Pulugurtha et al., 2013; Cai et al., 2016). The number of preschool, elementary and high school education facilities showed a negative relation to the number of all accidents and number of motorized mode accidents. This can be explained by the fact that in most school zones adequate measures are applied for reducing the drivers' intention to speed. In addition, parents were often together with users in segments near school gates, which made the drivers more cautious. Similar findings have been obtained in the city of Suzhou in China (Jia et al., 2018). This factor did not prove to have a statistically significant effect on the number of non-motorized mode accidents in the analyzed area, which suggests the further direction of research.

7. Conclusions

The study developed macro-level collision prediction models in traffic analysis zones in Novi Sad. In this paper, classical spatial regression models were developed to identify the spatial processes in motorized and non-motorized mode accidents. Moreover, demographic, traffic, road and environment characteristics which have an effect on the observed number of accidents were identified in this study.

The results of this research showed the presence of spatial autocorrelation in dependent variables which represent the total number of accidents and motorized mode accidents, so the spatial lag model proved to have the best modeling performances. With regard to non-motorized mode accidents, the spatial error model proved to have the best modeling performances, which confirms the assumption that spatial autocorrelation is present in residuals. In addition to evaluating the model, the factors that have an effect on the aggregate number of accidents in the area of traffic zones were determined in the paper at hand. Regarding the total number of accidents and motorized mode accidents, similar factors were identified. In addition to the factors that have an effect on the total number of accidents, motorized mode accidents are influenced by demographic characteristics which should be considered in future research. As far as non-motorized mode accidents are concerned, this type is affected by the characteristics related to vulnerable road users. The density of inhabitants, number of bus stops and length of cycleways are significant in non-motorized as opposed to motorized mode accidents.

The findings of this research can be of use to subjects who deal with traffic engineering, space planning and making decisions related to preparing certain activities necessary for road safety improvement in the analyzed area. The measures that need to be implemented pertain to the education of all categories of traffic users as well as the engineering modification of road infrastructure and road facilities. It is necessary to direct the education measures towards males since the percentage of males has a positive effect on traffic accidents. On the other hand, the engineering measures should be directed towards the activities regarding infrastructure improvement. The first approach needs to be directed towards building isolated pedestrian and cycling facilities and the second one needs to address equipping and protecting traffic facilities which are intended for public transport and regulating the parking service.

Apart from the analyzed factors, it is necessary to consider other macroscopic characteristics which lead to accidents in the observed area. The lack of other charactersitics is a limitation of this study. In order to include additional characteristics, it is necessary to invest more effort in terms of human resources, technology, finance and equipment, which is difficult to conduct in developing countries. Bearing this is mind, it can be concluded that this research makes a contribution to developing macroscopic analyses in specific areas, so that local communities could take measures to raise the level of road safety. Future studies should focus on examining significant factors in other areas as well as applying modern methods and models to get more reliable results.

References

AAA Foundation for Traffic Safety, 2016. Prevalence of Self-Reported Aggressive Driving Behavior: United States, 2014. AAA Foundation for Traffic Safety.

Abdel-Aty, M., Lee, J., Siddiqui, C., Choi, K., 2013. Geographical unit based analysis in the context of transportation safety planning. Trans. Res. Part A Policy Pract. 49, 62–75.

Abdel-Aty, M., Siddiqui, C., Huang, H., Wang, X., 2011. Integrating trip and roadway characteristics to manage safety in traffic analysis zones. Trans. Res. Rec. 2213 (1), 20–28.

Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury accidents in Pennsylvania. Accid. Anal. Prev. 38 (3), 618–625.

Álvarez, I.C., Barbero, J., Zofío, J.L., 2016. A spatial autoregressive panel model to analyze road network spillovers on production. Transp. Res. Part A Policy Pract. 93, 83–92.

Amoros, E., Martin, J.L., Laumon, B., 2003. Comparison of road accidents incidence and severity between some French counties. Accid. Anal. Prev. 35 (4), 537–547.

Anselin, L., 2001. Spatial Econometrics. A Companion to Theoretical Econometrics. 310330. .

Anselin, L., Griffith, D.A., 1988. Do spatial effecfs really matter in regression analysis? Papers Reg. Sci. 65 (1), 11–34.

Anselin, L., Bera, A.K., Florax, R., Yoon, M.J., 1996. Simple diagnostic tests for spatial dependence. Reg. Sci. Urban Econ. 26 (1), 77–104.

Anselin, L., Syabri, I., Kho, Y., 2006. GeoDa: an introduction to spatial data analysis. Geogr. Anal. 38 (1), 5–22.

Baass, K.G., 1981. Design of zonal systems for aggregate transportation planning models. Trans. Res. Rec. 807.

Black, W.R., Thomas, I., 1998. Accidents on Belgium's motorways: a network autocorrelation analysis. J. Transp. Geogr. 6.1, 23–31.

Cai, Q., Abdel-Aty, M., Lee, J., 2017b. Macro-level vulnerable road users crash analysis: a Bayesian joint modeling approach of frequency and proportion. Accid. Anal. Prev. 107, 11–19.

Cai, Q., Abdel-Aty, M., Lee, J., Eluru, N., 2017a. Comparative analysis of zonal systems for macro-level crash modeling. J. Saf. Res. 61, 157–166.

Cai, Q., Lee, J., Eluru, N., Abdel-Aty, M., 2016. Macro-level pedestrian and bicycle crash analysis: incorporating spatial spillover effects in dual state count models. Accid. Anal. Prev. 93, 14–22.

Chen, P., 2015. Built environment factors in explaining the automobile-involved bicycle crash frequencies: a spatial statistic approach. Saf. Sci. 79, 336–343.

Dale, M.R., Fortin, M.J., 2002. Spatial autocorrelation and statistical tests in ecology. Ecoscience 9 (2), 162–167.

Dong, N., Huang, H., Xu, P., Ding, Z., Wang, D., 2014. Evaluating spatial-proximity structures in crash prediction models at the level of traffic analysis zones. Trans. Res. Rec.: J. Trans. Res. Board 2432, 46–52.

Elvik, R., Bjørnskau, T., 2017. Safety-in-numbers: a systematic review and meta-analysis of evidence. Saf. Sci. 92, 274–282.

Elvik, R., Sørensen, M.W., Nævestad, T.O., 2013. Factors influencing safety in a sample of marked pedestrian crossings selected for safety inspections in the city of Oslo. Accid. Anal. Prev. 59, 64–70.

Fountas, G., Sarwar, M.T., Anastasopoulos, P.C., Blatt, A., Majka, K., 2018. Analysis of stationary and dynamic factors affecting highway accident occurrence: a dynamic correlated grouped random parameters binary logit approach. Accid. Anal. Prev. 113, 330–340.

Guo, Q., Xu, P., Pei, X., Wong, S.C., Yao, D., 2017. The effect of road network patterns on pedestrian safety: a zone-based Bayesian spatial modeling approach. Accid. Anal. Prev. 99, 114–124.

Guo, Y., Osama, A., Sayed, T., 2018. A cross-comparison of different techniques for modeling macro-level cyclist crashes. Accid. Anal. Prev. 113, 38–46.

Hadayeghi, A., Shalaby, A.S., Persaud, B.N., Cheung, C., 2006. Temporal transferability and updating of zonal level accident prediction models. Accid. Anal. Prev. 38 (3), 579–589.

Hakamies-Blomqvist, L., 1993. Fatal accidents of older drivers. Accid. Anal. Prev. 25 (1), 19–27.

Hong, J., Shankar, V.N., Venkataraman, N., 2016. A spatially autoregressive and heteroskedastic space-time pedestrian exposure modeling framework with spatial lags and endogenous network topologies. Anal. Methods Accid. Res. 10, 26–46.

Huang, H., Abdel-Aty, M., 2010. Multilevel data and Bayesian analysis in traffic safety. Accid. Anal. Prev. 42 (6), 1556–1565.

Huang, H., Abdel-Aty, M., Darwiche, A., 2010b. County-level crash risk analysis in Florida: bayesian spatial modeling. Trans. Res. Rec.: J. Trans. Res. Board 2148, 27–37.

Huang, H., Song, B., Xu, P., Zeng, Q., Lee, J., Abdel-Aty, M., 2016. Macro and micro models for zonal crash prediction with application in hot zones identification. J. Transp. Geogr. 54, 248–256.

Huang, Y., Wang, X., Patton, D., 2018. Examining spatial relationships between crashes and the built environment: a geographically weighted regression approach. J. Transp. Geogr. 69, 221–233.

Jacobsen, P.L., 2015. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. Inj. Prev. 21 (4), 271–275.

Janstrup Kira, H., 2017. Road Safety Annual Report 2017. OECD Publishing.

Jia, R., Khadka, A., Kim, I., 2018. Traffic crash analysis with point-of-interest spatial clustering. Accid. Anal. Prev. 121, 223–230.

Karlaftis, M.G., Tarko, A.P., 1998. Heterogeneity considerations in accident modeling. Accid. Anal. Prev. 30 (4), 425–433.

Kamel, M.B., Sayed, T., 2019. Accounting for mediation in cyclist-vehicle crash models: a Bayesian mediation analysis approach. Accid. Anal. Prev. 131, 122–130.

Koppel, S., Stephens, A.N., Bedard, M., Charlton, J.L., Darzins, P., Di Stefano, M., et al., 2019. Self-reported violations, errors and lapses for older drivers: measuring the change in frequency of aberrant driving behaviours across five time-points. Accid. Anal. Prev. 123, 132–139.

Ladron de Guevara, F., Washington, S., Oh, J., 2004. Forecasting accidents at the planning level: simultaneous negative binomial crash model applied in Tucson, Arizona. Transp. Res. Rec.: J. Trans. Res. Board 1897, 191–199.

Leden, L., 2002. Pedestrian risk decrease with pedestrian flow. A case study based on data from signalized intersections in Hamilton, Ontario. Accid. Anal. Prev. 34 (4), 457–464.

Lee, J., 2014b. Development of Traffic Safety Zones and Integrating Macroscopic and Microscopic Safety Data Analytics for Novel Hot Zone Identification.

Lee, J., Abdel-Aty, M., Jiang, X., 2014a. Development of zone system for macro-level traffic safety analysis. J. Transp. Geogr. 38, 13–21.

Lee, J., Abdel-Aty, M., Jiang, X., 2015. Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. Accid. Anal. Prev. 78, 146–154.

Lee, J., Abdel-Aty, M., Xu, P., Gong, Y., 2019. Is the safety-in-numbers effect still observed in areas with low pedestrian activities? A case study of a suburban area in the United States. Accid. Anal. Prev. 125, 116–123.

Lee, J., Yasmin, S., Eluru, N., Abdel-Aty, M., Cai, Q., 2018. Analysis of crash proportion by vehicle type at traffic analysis zone level: a mixed fractional split multinomial logit modeling approach with spatial effects. Accid. Anal. Prev. 111, 12–22.

LeSage, J.P., 1999. Spatial econometrics. Morgantown, WV.: Regional Research Institute. West Virginia University.

Levine, N., Kim, K.E., Nitz, L.H., 1995. Spatial analysis of Honolulu motor vehicle accidents: II. Zonal generators. Accid. Anal. Prev. 27 (5), 675–685.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Trans. Res. Part A Policy Pract. 44 (5), 291–305.

Lord, D., Miranda-Moreno, L.F., 2008. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle accidents: a Bayesian perspective. Saf. Sci. 46 (5), 751–770.

Loukaitou-Sideris, A., Liggett, R., Sung, H.G., 2007. Death on the crosswalk: a study of pedestrian-automobile collisions in Los Angeles. J. Plan. Educ. Res. 26 (3), 338–351.

Lusk, A.C., Furth, P.G., Morency, P., Miranda-Moreno, L.F., Willett, W.C., Dennerlein, J.T., 2011. Risk of injury for bicycling on cycle tracks versus in the street. Inj. Prev. 17 (2), 131–135.
MacNab C., Y., 2004. Bayesian spatial and ecological models for small-area accident and injury analysis. Accid. Anal. Prev. 36 (6), 1019–1028.
Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: methodological frontier and future directions. Anal. Methods Accid. Res. 1, 1–22.
Miaou, S.P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. Accid. Anal. Prev. 25 (6), 689–709.
Moran, P.A., 1948. The interpretation of statistical maps. J. R. Stat. Soc. Ser. B 10 (2), 243–251.
Ng, K.S., Hung, W.T., Wong, W.G., 2002. An algorithm for assessing the risk of traffic accident. J. Safety Res. 33 (3), 387–410.
Noland, R.B., Oh, L., 2004. The effect of infrastructure and demographic change on traffic-related fatalities and accidents: a case study of Illinois county-level data. Accid. Anal. Prev. 36 (4), 525–532.
Osama, A., Sayed, T., 2016. Evaluating the impact of bike network indicators on cyclist safety using macro-level collision prediction models. Accid. Anal. Prev. 97, 28–37.
Poch, M., Mannering, F., 1996. Negative binomial analysis of intersection-accident frequencies. J. Transp. Eng. 122 (2), 105–113.
Pulugurtha, S.S., Duddu, V.R., Kotagiri, Y., 2013. Traffic analysis zone level crash estimation models based on land use characteristics. Accid. Anal. Prev. 50, 678–687.
Quddus, M.A., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. Accid. Anal. Prev. 40 (4), 1486–1497.
Rhee, K.A., Kim, J.K., Lee, Y.I., Ulfarsson, G.F., 2016. Spatial regression analysis of traffic accidents in Seoul. Accid. Anal. Prev. 91, 190–199.
Saha, D., Alluri, P., Gan, A., Wu, W., 2018. Spatial analysis of macro-level bicycle accidents using the class of conditional autoregressive models. Accid. Anal. Prev.
Santamariña-Rubio, E., Pérez, K., Olabarria, M., Novoa, A.M., 2014. Gender differences in road traffic injury rate using time travelled as a measure of exposure. Accid. Anal. Prev. 65, 1–7.
Sawalha, Z., Sayed, T., 2001. Evaluating safety of urban arterial roadways. J. Transp. Eng. 127 (2), 151–158.
Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle accidents. Accid. Anal. Prev. 45, 382–391.
Strauss, J., Miranda-Moreno, L.F., Morency, P., 2013. Cyclist activity and injury risk analysis at signalized intersections: a Bayesian modelling approach. Accid. Anal. Prev. 59, 9–17.
Tasic, I., Elvik, R., Brewer, S., 2017. Exploring the safety in numbers effect for vulnerable road users on a macroscopic scale. Accid. Anal. Prev. 109, 36–46.
Ukkusuri, S., Hasan, S., Aziz, H., 2011. Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency. Trans. Res. Rec.: J. Trans. Res. Board (2237), 98–106.
Venkataraman, N., Shankar, V., Ulfarsson, G.F., Deptuch, D., 2014. A heterogeneity-in-means count model for evaluating the effects of interchange type on heterogeneous influences of interstate geometrics on crash frequencies. Anal. Methods Accid. Res. 2, 12–20.
Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Economet.: J. Economet. Soc. 307–333.
Wang, X., Jin, Y., Abdel-Aty, M., Tremont, P.J., Chen, X., 2012. Macro level model development for safety assessment of road network structures. Trans. Res. Rec. 2280, 100–109.
Wang, X., Yang, J., Lee, C., Ji, Z., You, S., 2016. Macro-level safety analysis of pedestrian accidents in Shanghai, China. Accid. Anal. Prev. 96, 12–21.
Wang, X., Zhou, Q., Yang, J., You, S., Song, Y., Xue, M., 2019. Macro-level traffic safety analysis in Shanghai, China. Accid. Anal. Prev. 125, 249–256.
Wang, Y., Kockelman, K.M., 2013. A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. Accid. Anal. Prev. 60, 71–84.
Wei, F., Lovegrove, G., 2013. An empirical tool to evaluate the safety of cyclists: community based, macro-level collision prediction models using negative binomial regression. Accid. Anal. Prev. 61, 129–137.
Wier, M., Weintraub, J., Humphreys, E.H., Seto, E., Bhatia, R., 2009. An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. Accid. Anal. Prev. 41 (1), 137–145.
World Health Organization, 2015. Global Status Report on Road Safety 2015. World Health Organization.
Wu, Y., 2014. A Comparative Analysis of Different Dilemma Zone Countermeasures at Signalized Intersections Based on Cellular Automaton Model.
Xie, K., Ozbay, K., Yang, H., 2019. A multivariate spatial approach to model crash counts by injury severity. Accid. Anal. Prev. 122, 189–198.
Xie, S.Q., Dong, N., Wong, S.C., Huang, H., Xu, P., 2018. Bayesian approach to model pedestrian accidents at signalized intersections with measurement errors in exposure. Accid. Anal. Prev. 121, 285–294.
Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: random parameter versus geographically weighting. Accid. Anal. Prev. 75, 16–25.
Xu, P., Huang, H., Dong, N., Abdel-Aty, M., 2014. Sensitivity analysis in the context of regional safety modeling: identifying and assessing the modifiable areal unit problem. Accid. Anal. Prev. 70, 110–120.
Yao, S., Loo, B.P., Lam, W.W., 2015. Measures of activity-based pedestrian exposure to the risk of vehicle-pedestrian collisions: space-time path vs. Potential path tree methods. Accid. Anal. Prev. 75, 320–332.
Zeng, Q., Huang, H., 2014. Bayesian spatial joint modeling of traffic accidents on an urban road network. Accid. Anal. Prev. 67, 105–112.
Zhai, X., Huang, H., Gao, M., Dong, N., Sze, N.N., 2018. Boundary crash data assignment in zonal safety analysis: an iterative approach based on data augmentation and Bayesian spatial model. Accid. Anal. Prev. 121, 231–237.
Zhou, Y., Wang, X., Holguín-Veras, J., 2016. Discrete choice with spatial correlation: a spatial autoregressive binary probit model with endogenous weight matrix (SARBP-EWM). Trans. Res. Part B Methodol. 94, 440–455.