

Multinomial logistic regression for prediction of vulnerable road users risk injuries based on spatial and temporal assessment

Mariana Vilaça^{a*}; Eloísa Macedo^a; Pavlos Tafidis^{a,1} and Margarida C. Coelho^a

mvilaca@ua.pt; macedo@ua.pt; pavlostafidis@ua.pt; margarida.coelho@ua.pt

ORCID: 0000-0002-0839-6869; 0000-0003-1503-1718; 0000-0002-8428-2345; 0000-0003-3312-191X

^a Centre of Mechanical Technology and Automation, Department of Mechanical Engineering, University of Aveiro, Campus Universitário de Santiago, 3810-193, Aveiro, Portugal

*corresponding author

¹ Current Affiliation: Construction Engineering Research Group, Faculty of Engineering Technology, University of Hasselt University, Agoralaan, B-3590 Diepenbeek, Hasselt, Belgium. pavlostafidis@uhasselt.be

Multinomial logistic regression for prediction of vulnerable road users risk injuries based on spatial and temporal assessment

Urban areas rapid growth often leads to negative effects such as traffic congestion and increasing accident risks due to the expansion in transportation systems. In the frame of smart cities, active modes are expected to be promoted to improve living conditions. To achieve this goal, it is necessary to reduce the number of vulnerable road users (VRUs) injuries. Considering injury severity levels from crashes involving VRUs, this paper seeks spatial and temporal patterns between cities and presents a model to predict the likelihood of VRUs to be involved in a crash. Kernel Density Estimation was applied to identify blackspots based on injury severity levels. A Multinomial Logistic Regression model was developed to identify statistically significant variables to predict the occurrence of these crashes. Results show that target spatial and temporal variables influence the number and severity of crashes involving VRUs. This approach can help to enhance road safety policies.

Keywords: road crashes, injury severity; kernel density estimation, multinomial logistic regression, vulnerable road users.

Introduction

Road crashes have a huge impact in society representing approximately 3% of the gross domestic product. Worldwide, 1.35 million people lost their lives every year in road crashes and between 20 and 50 million more people suffered non-fatal injuries (WHO, 2018). In 2015, 21% of the fatalities in European Union's roads were pedestrians, while 8% were cyclists (European Commission, 2016). In Portugal, in 2017, there were 5096 pedestrians and 1918 cyclists injured representing 20% and 5% of the road fatalities, respectively (ANSR, 2017).

The strategic target for EU road safety for the period of 2011–2020 is to reduce the number of road deaths by half (European Commission, 2011). Vulnerable Road Users (VRUs) such as pedestrians and cyclists suffer severe consequences in collisions since they are unprotected (European Commission, 2018). In order to reduce the number of VRU crashes or, at least, the severity of injuries, besides different improvement solutions

on redesign of infrastructures, solutions for mobility in the framework of smart cities can be developed (such as information tools to improve safety by, for instance, identifying areas prone to risk).

It is founded that cities that invest in active modes, such as walking and cycling, are reducing traffic congestion, which consequently makes travel times more reliable, reduced delays, crashes, increased access to city facilities and services and reduced transportation costs (Alliance for Biking & Walking, 2014). Ensuring VRUs road safety is also a way of promoting active transport modes, representing health, environment and economic positive effects.

Several studies have investigated spatial (Dereli & Erdogan, 2017; Jia, Khadka, & Kim, 2018; Mohaymany, Shahri, & Mirbagheri, 2013; Soltani & Askari, 2017; Van Raemdonck & Macharis, 2014; K. Xie, Ozbay, & Yang, 2019) and temporal patterns (Weast, 2018), or even a combination of both (Bao, Liu, & Ukkusuri, 2019; Liu & Sharma, 2018; Ma, Chen, & Chen, 2017) in the analysis of road crashes. The recognition of risky areas – often called blackspots - is the initial step in traffic safety analysis. A crash blackspot can be theoretically defined as any location that has a crash frequency significantly higher when compared to other areas (Van Raemdonck & Macharis, 2014). The spatial and temporal analysis of road crashes can also take into account the severity of the crash or the injury severity level. It was reported in the literature that considering the level of injury severity allow to avoid potential statistical problems and can change the idea of what could be a dangerous road zone (Mannering & Bhat, 2014). Additionally it is also important to point out that risk areas for crashes involving motor vehicles may present different characteristics when compared to areas of greatest risk for VRUs (Wang, Huang, & Zeng, 2017)

Recently, identification of spatial and temporal patterns among road crashes involving VRUs has become a hot research topic (Chimba, Musinguzi, & Kidando, 2018; Dozza, 2017; Loidl, Traun, & Wallentin, 2016; Lu, Mondschein, Buehler, & Hankey, 2018; Wang et al., 2017). This can be due to the fact that the number of VRUs fatalities and serious injuries have been growing, representing a challenge for both research and policymaking (Tiwari, 2018).

Regarding pedestrian crash patterns a study showed that the probability of severe injuries grows for older pedestrians, in males, rural areas, low-speed zones and with poor lighting (Senserrick, Boufous, de Rome, Ivers, & Stevenson, 2014). On the other hand, shopping and residential areas, pedestrians density is related with a reduction in pedestrian injury severity (Prato, Kaplan, Patrier, & Rasmussen, 2018). Regarding cyclists, studies revealed that urban roads and signal intersections density increase a crash risk (Guo, Osama, & Sayed, 2018). Likewise, presence of retail or service establishments, touristic attraction places, and environmental factors (e.g., time information, pavement condition and weather) increase the risk of vehicles-bicycle collisions (Prati, Pietrantonio, & Fraboni, 2017). On the other hand, posted speed limit and older age of the cyclists are related with increase of injury severity (Chen & Shen, 2016). Temporal correlations of crash reports revealed that pedestrian's fatalities occur specially in holiday periods and in November and December, while most cyclist fatalities occur in Summer or early Fall (Weast, 2018).

In the literature, predictive models have been developed for estimating the likelihood of VRUs to be involved in a crash. Logistic regression models were developed to analyze significance of contributing factors of VRUs crashes (Damsere-Derry, Palk, & King, 2017; Useche, Montoro, Alonso, & Oviedo-Trespalacios, 2018; Yuan & Chen, 2017). Yuan & Chen, (2017) revealed that night-time, road intersections, older age and

vehicle high-speed increase crash severity between VRUs. A prediction model was developed based on series of intersection crash models for total, severe, pedestrians and cyclist crashes and showed that macro-variables are significant for a rigorous crash analysis (Lee, Abdel-Aty, & Cai, 2017). Multinomial Logistic Regression (MLR) models have also been developed and showed effectiveness of the MLR approach in crash severity modeling (Abdulhafedh, 2017).

The research contribution of this paper is to perform a spatial and temporal analysis of crashes involving VRUs considering severity of injuries, in order to establish some pattern between cities with different specificities. A crash prediction model is also developed to identify the risk factors that can influence the severity of a VRU when involved in a motor-vehicle crash. A database of pedestrian and bicycle crashes was evaluated, comparing cities with different population densities and areas. The predictive model and spatial analysis are macro-level based and the blackspots are built on the density and severity of injuries. This work is built based on three steps:

1. To evaluate and perform a spatial mapping of blackspots using GIS (geographic information system) techniques and statistical data analysis procedures on the study areas (taking into account level of injury severity);
2. To perform a temporal analysis using spider plots, which are often used to reflect the trend of influencing variables and to compare multidimensional patterns;
3. To formulate a crash prediction model based on Multinomial Logistic Regression to describe the probability of a crash involving a motor-vehicle and VRU. This is important not only to predict crash occurrences in specific blackspots, but also to infer the severity of crashes involving VRUs.

This work intends to be a baseline study supported by a thorough analysis that can be used by policy and decision makers and road safety managers in order to clearly recognize blackspots and address the most relevant variables that influence VRUs safety.

Methodology

In this chapter, a description of methods applied for blackspots identification and development of the predictive models is made. Afterwards, case studies and the development of crash database is described. The conceptual framework is presented in Figure 1.

Blackspots Identification

To highlight areas prone to road crashes involving VRUs, geostatistical-based approach KDE was applied to obtain patterns based on the level of VRUs injury severity using the ArcGIS® software (ERSI - Environmental Sistem Research Institute, 2015).

KDE is one of the most commonly used methods and revealed to outperform other popular methods for spatial analysis of crashes and blackspots identification (Yu, Liu, Chen, & Wang, 2014). In this technique, each observation is covered by a kernel, yielding a circular-cell-shape neighbourhood, with maximum value at a reference point, decreasing to zero at radius (r) distance from it.

In this study, we used the quartic kernel function (default in ArcGIS), which is one of the most commonly used functions. The density estimation using such function can be given as

$$f(x, y) = \frac{1}{mr^2} \sum_{i=1}^m K \left(\frac{d_i}{r} \right), \quad (1)$$

where $f(x, y)$ is the density estimation for location (x, y) , m is the number of observations, K is the kernel function defined as

$$K\left(\frac{d_i}{r}\right) = \begin{cases} K\left(1 - \frac{d_i^2}{r^2}\right), & 0 < d_i \leq r \\ 0, & d_i > r \end{cases}, \quad (2)$$

with d_i being the i -th observation location and K a real coefficient.

The choice of bandwidth r controls the smoothness of the estimated density (Z. Xie & Yan, 2008). Considering the data level of detail and the area of each city under study, different radius (bandwidths) were empirically examined and smaller radius showed to be reasonably suitable to obtain an unsmooth density distribution, allowing smaller scale details. In the present study, the radius was set to approximately 175 meters.

In order to embed the injury severity level in the data spatial structure, a specific weight on each VRU injured observation was considered based on the severity index developed by Elvik (2008). This approach establishes different weights for the different severity levels: 1 for light injuries, 3 for serious injuries and 5 for fatalities.

KDE returned nine levels of risk, displayed in a range of grey shades from nearly white (Level 9: low-risk area) to black (Level 1: high-risk area), and Kernel density surfaces were derived for total injuries for each city.

Predictive Modelling

MLR is a predictive analysis that is used to describe data and to explain the relationship between the dependent nominal variable and one or more independent variables. In this paper, the response variable is categorical; it has two classes related to VRUs: pedestrian or cyclist. The set of predictor variables includes VRUs' gender and age, level of injury severity, weekday, time period and weather conditions. By using MLR, one can determine the strength of influence that a particular independent variable has upon the type of VRU involved in a crash. We assume that the pedestrian is the reference group since it is the class with more injuries. The statistical software SPSS was used (IBM Corp., 2016).

Considering an MLR model in which the response variable consists of $k \geq 2$

categories, the probability of a given observation x belong to one of the groups y_i can be estimated by

$$P(Y = y_i|x) = \frac{\exp\{\beta_{i0} + \beta_{i1}x_1 + \beta_{i2}x_2 + \dots + \beta_{in}x_n\}}{1 + \sum_{i=1}^{k-1} \exp\{\beta_{i0} + \beta_{i1}x_1 + \beta_{i2}x_2 + \dots + \beta_{in}x_n\}}, \quad (4)$$

where $i = 1, 2, \dots, k - 1$, x_i is the i -th independent variable of the data set, and β_i represent the estimated model. In particular, the MLR was performed considering a 95% confidence interval. Estimation of the parameters of these models was conducted using maximum likelihood procedures. The well-known Deviance and Pearson chi-square tests were used as goodness-to-fit statistics to evaluate the model fit. The Pseudo- R^2 statistic was used as a measure on how well the model can predict the dependent variable based on the independent variables. The used methods for computing this measure were Cox and Snell, Nagelkerke and McFadden, which are most often available in statistical software. Finally, a likelihood ratio test was performed for evaluating the effect of each of the parameters, providing the weight of each independent variable in the prediction model.

Development of a crash database

Crash data involving VRUs from three Portuguese cities (Aveiro, Porto and Lisbon) with different areas and socio-demographic characteristics (Table 1) were analysed. Furthermore, a crash database was conceived. A total of 4439 VRUs-involving crash registrations from 2012-2015 were provided by ANSR (Portuguese Authority of Road Safety). 4615 VRUs were injured in these crashes. 87% of the injuries are related to motor vehicle-pedestrians and 13% to motor vehicle-cyclists crashes. All cities present a percentage of light injuries between 90 to 97% of total injuries. Table 2 describes the distribution of number of injuries per 10000 inhabitants and per square kilometre for each city.

The crashes database was built and the analysis was focused on the VRU injury severity level, which is subdivided into three classes: light injuries, serious injuries, and fatalities. According to ANSR, a victim is considered seriously injured if there is the need to be at least 24 hours in a hospital but is discharged within 30 days after the crash; fatalities involve victims that do not survive within those 30 days (ANSR, 2017).

The main attributes considered in the forthcoming analyses are:

- VRU age and injury severity level;
- Temporal variables: year, month, weekday, time of the day;
- Weather conditions: good, bad (adverse weather conditions, e.g., rain, fog, snow).

A more specific analysis is developed focusing on the most severe consequences (severe injuries and fatalities) in order to identify patterns between them. This analysis considered the attributes mentioned above and a detailed look is given to the following ones:

- VRU Gender (male, female) and presence of the most vulnerable age groups (14 or younger, and 65 or older);
- Type of road location (segment, intersection and others);
- Built environment (area characterization: shopping, touristic, educational, health, residential, industrial, services and agriculture).

Results

This section presents the results of spatial and temporal analysis, in an attempt to discover spatial and temporal patterns among the severely injured and fatal VRUs. Finally, results on multinomial logit models are discussed.

Spatial Analysis

Crashes involving VRUs were georeferenced and an injury attribute was used to generate

spatial maps. Figure 2 illustrates the geographic distribution of crashes resulting on VRUs injuries by level of severity, highlighting blackspots on each city. KDE was applied to analyse spatial distribution of motor-vehicle-VRU crashes.

Regarding Aveiro, the main blackspot is inside the city centre, including a shopping area and one of the main inner-city connection roads. Porto blackspots were identified in urban and historical centres, in places involving high number of tourist points, churches, stores and train station. Lisbon blackspots are mostly in urban and historical centres, close to touristic points, but also around governmental institutions. Additionally, in a second level, there is a blackspot covering a train station and a hospital. Results suggest blackspots in areas that attract many people and although it can be thought as expected, the truth is that this is a compelling situation, since in these specific areas vehicles running speed should be low (speed limit equal or less than 50km/h).

Temporal Analysis

Considering an annual evolution, Aveiro is the only city presenting a decrease of 2% regarding pedestrians' injuries. Regarding cyclists, Aveiro presents an annual growth rate of 11%. Porto presents an annual growth rate of 2% of pedestrians' injuries and 15% of cyclists' injuries. Annual growth rate in Lisbon is 4% of pedestrians' and 15% of cyclists' injuries. A closer look regarding pedestrian's number of injuries shows Lisbon with an increase over the years and Aveiro and Porto with a decrease in 2015. Considering injured cyclists, the proportion of this class is higher in Aveiro, followed by Porto, and Lisbon. Decrease of cyclists injured in 2015 can be explained by the implementation of new legislation in 2014, the driver of the motor vehicle is required to leave a minimum lateral distance of 1.5 meters between the vehicle and the bicycle.

Figure 3 shows the monthly evolution of VRUs injured for the cities and years under study considering the level of injury severity (light and serious injuries and

fatalities). A first observation is that the light injuries represent a significant weight when compared to the other classes, and thus, contribute more to any analyses. Aveiro has the most evident fluctuation during the months of October and November. August, which typically is a vacation month, presented lower number of injuries, particularly for Porto and Lisbon. Regarding Porto, VRUs injuries' peak is in September, then decreasing until November. Concerning Lisbon data, the highest number of VRUs injuries occur in December and October. Considering a global analysis of the three cities, November, October and September are the months with higher number of VRUs injuries with around 1,5 injury/1000 inhabitants.

Table 3 presents the distribution of VRU injuries for weekday. Regarding Aveiro, 20% of VRUs injuries occur on Thursdays and 18% during weekend. In Porto, weekends still reach lower percentage of injuries, while the riskiest day seems to be Friday with 18% of injuries. Focusing on Lisbon, crashes involving VRUs injuries occur mostly on Thursdays (18%) and Fridays (17%), and weekends represent 17% of injuries. An overview of the three cities allow us to conclude that Thursday and Friday are the weekdays with more injuries (35% of the total; 2,7 and 2,5 injuries/1000 inhabitants, respectively).

Figure 4 shows distribution of vehicle-pedestrian and -cyclist crashes for different hours, during weekdays. Spider plots for Aveiro suggest most vehicle-pedestrian crashes occur on Monday with peak hours at 8am and 5pm, while for cyclists most critical days are Thursday, at 1pm, and Mondays at 8am. For pedestrians, peak hours for Porto seems to be 6pm on Mondays, 5pm and 7pm on Fridays, and 9am and 6pm on Wednesdays. Regarding cyclists, peaks are clearly at 7pm on Thursdays and there is also a peak at 4pm on Wednesdays. In Lisbon, number of weekend pedestrians-involving crashes is smaller when compared to other days. Typically, Thursdays have peak hours in terms of crashes

between 8am and 9am, and 5pm to 6pm, while on Wednesdays, peaks are at 9am and 11am, and between 5pm and 6pm. For cyclists, Thursday at 9am and 7pm, Saturday at 11am, Tuesday at 4pm and Monday at 6pm represent the most critical time periods. Results of the three cities revealed that the worst hours considering pedestrians injuries are between 5pm to 7pm and for cyclist at 1pm and, as for pedestrians, from 5pm to 7pm. This is explained by VRUs daily routines and peak hours with higher traffic volumes.

Figure 5 presents distribution of injuries according to different age groups. Specifically, in Aveiro, school age group of pedestrians (<15 yr. old) reveals a peak in crashes at 8am and most crashes occur after 4pm for pedestrians between 18 and 49 years of age. For pedestrians with 65+ years, 8am and 10-11am are critical periods. Cyclists between 25 and 49 years are involved in a higher number of crashes at 10am, 1pm and 7pm. For older cyclists, 11am is a critical hour. Patterns for pedestrian-involving crashes can be visualized at 8am for school age, between 5pm and 8pm for 18-24 and 25-49 age groups, during morning for 65+. Part of the reason of these results may be due to existence of schools, University and hospital close to city center, which involve many daily trips. Comparing to Aveiro, distinct patterns can be pointed in Porto. Pedestrians of 65+ clearly presented the biggest number of injuries and have a first high-risk time at 9am, as well as the working age group (25-64). However, other critical times for older pedestrians are 11am and 6pm, while the working age group has a peak between 5pm and 7pm. Many vehicle-cyclist crashes involve working age group, with peaks between 9am and 11am, 3pm and 4pm, and also at 7pm. These findings suggest off-peak traffic hours are also important in crashes involving older pedestrians, since their daily routines are not restricted to working hours. In Lisbon, most of crashes involve pedestrians on working age group and 65+ groups, with morning peaks between 8am and 9am, and 8am and 11am, respectively. During afternoon, working age group has peaks in terms of crashes

around lunch time and between 4 to 8pm, and older pedestrians present a peak at 5pm. For cyclists between 25 and 49 years of age, there are more crashes at 8am and 12am, as well as between 4-7pm, which can be explained by the existence of schools, general services and offices in city center.

Analysis of Severely Injured and Fatal occurrences

A closer look to severely injured and fatal occurrences is given considering the importance to find patterns between the most severe consequences.

Aveiro presents a higher proportion in the number of serious injuries and fatalities (11% considering pedestrians' injuries and 9% considering cyclists injuries). Regarding Porto and Lisbon these percentages are for pedestrians' injuries 3% and 7%, respectively. Regarding cyclists injuries Porto did not present any serious or fatal crash and Lisbon presented a percentage of 5%.

Month evolution (Figure 3) revealed that May and February are the months with highest serious injuries and fatalities, respectively, for the city of Aveiro. For Porto, February has the highest number of serious injuries, while the months of September, November and December present more fatalities. In Lisbon, serious injuries occur in May and January recorded the highest number of fatalities. As a general overview, January and February seem to be the months with more fatalities among all cities.

Regarding the distribution of VRU injuries for weekdays (Table 3) Aveiro present the most severe injuries (serious and fatal) on Mondays. For Porto, Wednesday and Thursday present the huge percentage of serious injuries and fatalities (4%). Lisbon, Sunday present the highest percentage (9%) of severe injuries. A general overview highlight Thursday and Friday the most critical weekdays in what concern high VRUs injury severity levels.

Considering VRUs gender, results show that in Aveiro and Lisbon almost 80%

and 90%, respectively, are male cyclists. The trend concerning male pedestrians is more balanced in Porto and Lisbon, representing 54% and 57%, respectively, while in Aveiro more than 70% of the injuries occur with female pedestrians.

Regarding the most vulnerable age groups, results show that 7% of the injured pedestrians in Porto and Lisbon are children, while in Aveiro this percentage raises up to 28%. Half of the elderly pedestrians involved in crashes in Porto are severely injured or fatalities, while in Lisbon such value drops to 38%. Aveiro presented the smallest percentage (17%). Considering cyclists, Aveiro presented a massive 43% of elderly cyclists between the severely injured and fatalities, while Lisbon present 6%. Aveiro does not present any severely injured or dead child cyclist, while Lisbon presents almost 20% cyclists severely injured or dead.

With respect to type of road locations, results show that, for serious injuries and fatalities of pedestrians for all cities, 33-65% occur in segments and 28-41% in intersections. About cyclists, results revealed that around 42% occur in segments and 49% in intersections. A residual percentage of the severe and fatal injuries occur in roundabouts for all cities.

The following spider plots illustrate the temporal distribution of severely injured and dead VRUs considering different road types: straight segments, roundabouts, intersections, or other road location (parking lot road, private road, etc.) – Figures 6 and 7.

A closer look on some road specificities allows to conclude that more than 40% of the pedestrian-motor vehicle crashes occur in the presence of crosswalks; in particular, Porto presented the worst scenario with almost 60%. Porto presents the highest percentage of crashes occurring in the presence of traffic lights (almost 40%), followed by Lisbon with 25%, while Aveiro presented the smallest percentage (3%). None of the

records involving severely injured and dead cyclists describe occurrences close to cycle lanes, however 8% of the total crashes involving light injured cyclists occurred close to these facilities.

Figure 8 illustrates the relative proportion of severely injured and dead VRUs for each particular city taking into account aspects of built environment. Regarding built environment, results show that Aveiro and Lisbon present the highest number in residential areas (34% and 40%, respectively). There are also relevant percentages of crashes in agricultural and industrial areas in Aveiro (19% and 16%, respectively); this can be explained by the higher speed limit (in the roads close to industrial areas) and the lack of sidewalks in agricultural areas. On the other hand, Lisbon areas has a higher exposure of VRUs exposure in touristic (23%) and service (17%) areas. Porto presents a quite different daunting trend with 28% of the most severe crashes occurring in educational areas, 24% in residential areas, and 15% in touristic points. Shopping areas can be associated to 6-8% of the severely injured and dead VRUs for all cities.

Multivariate Model Analysis

In this section, MLR models involving vehicle-VRU crashes for each city are presented. Table 4 presents the overall statistical significance values of each developed model.

Results suggest appropriate fits to the models and show that the variables added to the model are statistically significantly (Sig.<0.05) and improve the model for each city (i.e., the obtained models significantly predict the response variable). Results on the quality of fitting the data (goodness-to-fit) reveal that the model fits well the data for Porto and Lisbon, while for Aveiro the results of both measures of goodness-to-fit are different, yet Pearson chi-square statistic shows the model fits the data as well. The pseudo-R² statistics permits to assess the predictive strength of the obtained MLR models. The best measure is obtained for the Nagelkerke's R²: 37%, 39% and 29% for

Aveiro, Porto and Lisbon, respectively, meaning that the obtained models are able to explain these percentages of data variability. Finally, regarding statistics related to the model parameters, VRU gender, age group, and weather conditions are statistically significant for all models.

Conclusion

This paper presented a spatial and temporal analysis of crashes involving motor-vehicles and VRUs considering severity of injuries, in an attempt to highlight some patterns between cities of different specificities. Moreover, assessing which factors can influence the level of injury severity of a VRU was also reported.

The main findings allow to conclude that most injuries occur in surrounding areas of high attraction places, such as train stations, shopping and touristic points, where speed limits are relatively low. Intersections are the singularity type with more serious injuries and fatal cyclists, while road segments seem to be more dangerous for pedestrians. More than 40% of the pedestrian crashes occur in the vicinity of crosswalks. A general overview on the built environment allows to conclude that areas with more impact are residential, educational and touristic zones. In a medium-sized city, as Aveiro, pedestrians from the active age and female groups are the most vulnerable. However, for Porto and Lisbon, elderly people are the most vulnerable both in injury number and in severity; in these cities, cyclists in the active age are more likely to be involved in a crash. The developed MLR models for each city revealed that VRU gender and age, as well as weather conditions, are statistically significant variables to predict this type of crashes.

Despite the findings achieved so far, some limitations should be taken into account and addressed in future research. First, data of pedestrians and cyclists' exposure by age and gender could give a better perspective of the result. Secondly, a micro level

study can be attempted with additional information related to vehicle details, road characteristics and driver profile information.

This work intends to be useful for policy and decision makers, as well as road safety managers, in order to clearly recognize blackspots and improve VRUs safety. This is even more important in an era where driverless vehicles are about to be implemented, and the way they will circulate in the urban space and interact with VRUs is of utmost important.

Acknowledgments

The authors acknowledge ANSR for providing crashes data and the support of TEMA – CENTRO 01-0145-FEDER-022083; Strategical Project UID/EMS/00481/2013 (FCT-Portuguese Science and Technology Foundation); @CRUiSE project (PTDC/EMS-TRA/0383/2014), funded within Project 9471 – Reforçar a Investigação, o Desenvolvimento Tecnológico e a Inovação and supported by European Community Fund FEDER; MobiWise (P2020 SAICTPAC/0011/2015), co-funded by COMPETE2020, Portugal2020 - Operational Program for Competitiveness and Internationalization (POCI), European Union's ERDF (European Regional Development Fund), and FCT; and CISMOB (PGI01611, funded by Interreg Europe Programme). This work was also financially supported by the project POCI-01-0145-FEDER-029463 (DICA-VE) and POCI-01-0145-FEDER-029679 (InFLOWence) funded by FEDER through COMPETE2020- Programa Operacional Competitividade e Internacionalização (POCI), and by national funds (OE), through FCT/MCTES. Finally, the authors acknowledge the collaboration of the students Laura Laranjo and Bárbara Romeira for the collaboration in data georeferencing.

References

- Abdulhafedh, A. (2017). Incorporating the Multinomial Logistic Regression in Vehicle Crash Severity Modeling: A Detailed Overview. *Journal of Transportation Technologies*, 7(3), 279–303. <https://doi.org/10.4236/jtts.2017.73019>
- Alliance for Biking & Walking. (2014). Bicycling and walking in the United States: 2012 Benchmarking Report, 1–248. Retrieved from <http://www.bikewalkalliance.org/download-the-2014-benchmarking-report%5Cnhttp://www.peoplepoweredmovement.org>
- ANSR. (2017). *Annual Report 2017 (30 days victims) - National Authority of Road Safety [in Portuguese]*. Portuguese National Authority for Road Safety.
- Bao, J., Liu, P., & Ukkusuri, S. V. (2019). A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data. *Accident Analysis & Prevention*, 122, 239–254. <https://doi.org/10.1016/J.AAP.2018.10.015>
- Chen, P., & Shen, Q. (2016). Built environment effects on cyclist injury severity in automobile-involved bicycle crashes. *Accident Analysis and Prevention*, 86, 239–246. <https://doi.org/10.1016/j.aap.2015.11.002>
- Chimba, D., Musinguzi, A., & Kidando, E. (2018). Associating pedestrians crashes with demographic and socioeconomic factor. *Case Studies on Transport Policy*, 6(January), 11–16.
- Comunidade Intermunicipal da Região de Aveiro (CIRA). (2012). *Plano Municipal de Mobilidade de Aveiro [in portuguese]*. Aveiro.
- Damsere-Derry, J., Palk, G., & King, M. (2017). Road accident fatality risks for “vulnerable” versus “protected” road users in northern Ghana. *Traffic Injury Prevention*, 18(7), 736–743. <https://doi.org/10.1080/15389588.2017.1302083>
- Dereli, M. A., & Erdogan, S. (2017). A new model for determining the traffic accident black spots using GIS-aided spatial statistical methods. *Transportation Research Part A: Policy and Practice*, 103, 106–117. <https://doi.org/10.1016/j.tra.2017.05.031>
- Dozza, M. (2017). Crash risk: How cycling flow can help explain crash data. *Accident Analysis & Prevention*, 105, 21–29. <https://doi.org/10.1016/J.AAP.2016.04.033>

- Elvik, R. (2008). *State-of-the-art approaches to road accident black spot management and safety analysis of road network*. TOI report 883/2007.
- ERSI - Environmental Sistem Research Institute. (2015). About ArcGIS | Mapping & Analytics Platform. Redlands, CA. Retrieved from <https://www.esri.com/en-us/arcgis/about-arcgis/overview>
- European Comission. (2011). White Paper on transport - Roadmap to a single European transport area - Towards a competitive and resource-efficient transport system. Luxembourg: Plublications Office of the European Union. <https://doi.org/10.2832/30955>
- European Commission. (2016). *Road Safety In The European Union - Trends, statistics and main Challenges*. 2363-4383 Brussels. <https://doi.org/10.2832/993015>
- European Commission. (2018). *Pedestrians and cyclists*. European Commission. Directorate General for Transport, February, 2018.
- Guo, Y., Osama, A., & Sayed, T. (2018). A cross-comparison of different techniques for modeling macro-level cyclist crashes. *Accident Analysis and Prevention*, 113, 38–46. <https://doi.org/10.1016/j.aap.2018.01.015>
- IBM Corp. (2016). IBM SPSS Software | IBM Analytics. New York. Retrieved from <https://www.ibm.com/analytics/data-science/predictive-analytics/spss-statistical-software>
- INE. (2011). *Censos 2011 Resultados Definitivos - Portugal [in portuguese]*. Portuguese Statistics Institute.
- INE. (2017). *Estatísticas dos Transportes e Comunicações 2016 [In Portuguese]*. Portuguese Statistics Institute.
- Jia, R., Khadka, A., & Kim, I. (2018). Traffic crash analysis with point-of-interest spatial clustering. *Accident Analysis and Prevention*, 121, 223–230. <https://doi.org/10.1016/j.aap.2018.09.018>
- Lee, J., Abdel-Aty, M., & Cai, Q. (2017). Intersection crash prediction modeling with macro-level data from various geographic units. *Accident Analysis and Prevention*, 102, 213–226. <https://doi.org/10.1016/j.aap.2017.03.009>
- Liu, C., & Sharma, A. (2018). Using the multivariate spatio-temporal Bayesian model to

- analyze traffic crashes by severity. *Analytic Methods in Accident Research*, 17, 14–31. <https://doi.org/10.1016/j.amar.2018.02.001>
- Loidl, M., Traun, C., & Wallentin, G. (2016). Spatial patterns and temporal dynamics of urban bicycle crashes-A case study from Salzburg (Austria). *Journal of Transport Geography*, 52, 38–50. <https://doi.org/10.1016/j.jtrangeo.2016.02.008>
- Lu, T., Mondschein, A., Buehler, R., & Hankey, S. (2018). Adding temporal information to direct-demand models: Hourly estimation of bicycle and pedestrian traffic in Blacksburg, VA. *Transportation Research Part D: Transport and Environment*, 63, 244–260. <https://doi.org/10.1016/j.trd.2018.05.011>
- Ma, X., Chen, S., & Chen, F. (2017). Multivariate space-time modeling of crash frequencies by injury severity levels. *Analytic Methods in Accident Research*, 15, 29–40. <https://doi.org/10.1016/j.amar.2017.06.001>
- Mannering, F. L., & Bhat, C. R. (2014). Analytic methods in accident research: Methodological frontier and future directions. *Analytic Methods in Accident Research*, 1, 1–22. <https://doi.org/10.1016/J.AMAR.2013.09.001>
- Mohaymany, A. S., Shahri, M., & Mirbagheri, B. (2013). GIS-based method for detecting high-crash-risk road segments using network kernel density estimation. *Geo-Spatial Information Science*, 16(2), 113–119. <https://doi.org/10.1080/10095020.2013.766396>
- Prati, G., Pietrantonio, L., & Fraboni, F. (2017). Using data mining techniques to predict the severity of bicycle crashes. *Accident Analysis and Prevention*, 101, 44–54. <https://doi.org/10.1016/j.aap.2017.01.008>
- Prato, C. G., Kaplan, S., Patrier, A., & Rasmussen, T. K. (2018). Considering built environment and spatial correlation in modeling pedestrian injury severity. *Traffic Injury Prevention*, 19(1), 88–93. <https://doi.org/10.1080/15389588.2017.1329535>
- Senserrick, T., Boufous, S., de Rome, L., Ivers, R., & Stevenson, M. (2014). Detailed Analysis of Pedestrian Casualty Collisions in Victoria, Australia. *Traffic Injury Prevention*, 15, S197–S205. <https://doi.org/10.1080/15389588.2014.935356>
- Soltani, A., & Askari, S. (2017). Exploring spatial autocorrelation of traffic crashes based on severity. *Injury*, 48(3), 637–647. <https://doi.org/10.1016/j.injury.2017.01.032>
- Tiwari, G. (2018). Safety of ‘the vulnerable road users’: current challenges and need for

- a new approach. *International Journal of Injury Control and Safety Promotion*, 25(1), 1–2. <https://doi.org/10.1080/17457300.2018.1429211>
- Useche, S., Montoro, L., Alonso, F., & Oviedo-Trespalacios, O. (2018). Infrastructural and Human Factors Affecting Safety Outcomes of Cyclists. *Sustainability*, 10(2), 299. <https://doi.org/10.3390/su10020299>
- Van Raemdonck, K., & Macharis, C. (2014). The Road Accident Analyzer: A Tool to Identify High-Risk Road Locations. *Journal of Transportation Safety and Security*, 6(2), 130–151. <https://doi.org/10.1080/19439962.2013.826314>
- Wang, J., Huang, H., & Zeng, Q. (2017). The effect of zonal factors in estimating crash risks by transportation modes: Motor vehicle, bicycle and pedestrian. *Accident Analysis and Prevention*, 98, 223–231. <https://doi.org/10.1016/j.aap.2016.10.018>
- Weast, R. (2018). Temporal factors in motor-vehicle crash deaths: Ten years later. *Journal of Safety Research*, 65, 125–131. <https://doi.org/10.1016/j.jsr.2018.02.011>
- WHO. (2018). *Global Status Report on Road Safety 2018*. Geneva: World Health Organization; 2018. Licence: CC BY-NC-SA 3.0 IGO.
- Xie, K., Ozbay, K., & Yang, H. (2019). A multivariate spatial approach to model crash counts by injury severity. *Accident Analysis & Prevention*, 122, 189–198. <https://doi.org/10.1016/J.AAP.2018.10.009>
- Xie, Z., & Yan, J. (2008). Kernel Density Estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems*, 32(5), 396–406. <https://doi.org/10.1016/j.compenvurbsys.2008.05.001>
- Yu, H., Liu, P., Chen, J., & Wang, H. (2014). Comparative analysis of the spatial analysis methods for hotspot identification. *Accident Analysis and Prevention*, 66, 80–88. <https://doi.org/10.1016/j.aap.2014.01.017>
- Yuan, Q., & Chen, H. (2017). Factor comparison of passenger-vehicle to vulnerable road user crashes in Beijing, China. *International Journal of Crashworthiness*, 22(3), 260–270. <https://doi.org/10.1080/13588265.2016.1248226>

TABLES

Table 1. Socio-demographic characteristics from the case studies cities [Comunidade Intermunicipal da Região de Aveiro (CIRA), 2012; INE, 2011, 2017].

| | Population (inhabitants) | Area (km²) | Percentage of trips (%) | | Road network by district (km) |
|---------------|-------------------------------------|----------------------------------|--------------------------------|----------------|--|
| | | | Walking | Cycling | |
| Aveiro | 78450 | 197.6 | 21.0 | 2.7 | 611 |
| Porto | 237591 | 41.4 | 21.6 | 0.2 | 896 |
| Lisbon | 547733 | 100.0 | 19.4 | 0.2 | 843 |

Table 2. Distribution of VRU injuries in the case studies.

| | Injuries per 10000 inhabitants | | Injuries per km ² | |
|---------------|--------------------------------|----------|------------------------------|----------|
| | Pedestrians | Cyclists | Pedestrians | Cyclists |
| Aveiro | 21 | 20 | 1 | 1 |
| Porto | 51 | 6 | 30 | 4 |
| Lisbon | 49 | 6 | 27 | 3 |

Table 3. Consequences of road crashes involving VRU, for each severity level, along the different days of the week

| | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday | Total | |
|---------------|-----------------------|---------|-----------|----------|--------|----------|--------|-------|-------------|
| Aveiro | Light Injury | 46 | 47 | 44 | 56 | 41 | 29 | 24 | 287 |
| | Serious Injury | 6 | 2 | 2 | 6 | 6 | 3 | 0 | 25 |
| | Fatal | 2 | 0 | 1 | 1 | 1 | 1 | 1 | 7 |
| | TOTAL | 54 | 49 | 47 | 63 | 48 | 33 | 25 | 319 |
| Porto | Light Injury | 213 | 181 | 220 | 219 | 230 | 133 | 91 | 1287 |
| | Serious Injury | 4 | 4 | 5 | 7 | 3 | 4 | 2 | 29 |
| | Fatal | 3 | 2 | 4 | 3 | 4 | 1 | 0 | 17 |
| | TOTAL | 220 | 187 | 229 | 229 | 237 | 138 | 93 | 1333 |
| Lisbon | Light Injury | 408 | 444 | 468 | 491 | 467 | 260 | 213 | 2751 |
| | Serious Injury | 21 | 29 | 26 | 28 | 36 | 17 | 18 | 175 |
| | Fatal | 5 | 7 | 4 | 8 | 6 | 3 | 4 | 37 |
| | TOTAL | 434 | 480 | 498 | 527 | 509 | 280 | 235 | 2963 |

Table 4. Overall statistical significance values for the independent variables.

a) **Model Fitting Information**

| Model | Model Fitting Criteria | | Likelihood Ratio Tests | | |
|---------------|------------------------|----------------|------------------------|----|-------|
| | -2 Log Likelihood | Log Likelihood | Chi-Square | df | Sig. |
| Aveiro | 278.105 | | 104.15 | 19 | 0.000 |
| Porto | 336.391 | | 261.812 | 19 | 0.000 |
| Lisbon | 681.761 | | 449.502 | 19 | 0.000 |

b) **Goodness-to-Fit**

| | | Chi-Square | df | Sig. |
|---------------|-----------------|------------|-----|-------|
| Aveiro | Pearson | 217.684 | 191 | 0.09 |
| | Deviance | 233.861 | 191 | 0.019 |
| Porto | Pearson | 463.113 | 439 | 0.206 |
| | Deviance | 225.366 | 439 | 1 |
| Lisbon | Pearson | 539.061 | 679 | 1 |
| | Deviance | 413.721 | 679 | 1 |

c) **Pseudo R Square**

| | Cox and Snell | Nagelkerke | McFadden |
|---------------|---------------|------------|----------|
| Aveiro | 0.279 | 0.371 | 0.236 |
| Porto | 0.178 | 0.385 | 0.316 |
| Lisbon | 0.141 | 0.293 | 0.231 |

d) **Likelihood Ratio Tests**

| Effect | Model Fitting Criteria | | Likelihood Ratio Tests | | |
|---------------|---------------------------|----------------|------------------------|----|-------|
| | -2 Log Likelihood | Log Likelihood | Chi-Square | df | Sig. |
| Aveiro | VRU Gender | 337.782 | 59.677 | 1 | 0 |
| | VRU Age Group | 297.152 | 19.047 | 5 | 0.002 |
| | Injury Severity | 280.222 | 2.117 | 2 | 0.347 |
| | Weekday | 290.388 | 12.283 | 6 | 0.056 |
| | Time Period | 285.086 | 6.981 | 4 | 0.137 |
| | Weather Conditions | 290.286 | 12.182 | 1 | 0 |
| | VRU Gender | 466.126 | 129.735 | 1 | 0 |
| | VRU Age Group | 408.275 | 71.883 | 5 | 0 |
| | Injury Severity | 342.571 | 6.18 | 2 | 0.045 |
| Porto | Weekday | 358.881 | 22.49 | 6 | 0.001 |
| | Time Period | 357.194 | 20.803 | 4 | 0 |
| | Weather Conditions | 340.832 | 4.44 | 1 | 0.035 |
| | VRU Gender | 918.833 | 237.072 | 1 | 0 |
| | VRU Age Group | 849.876 | 168.115 | 5 | 0 |
| Lisbon | Injury Severity | 685.963 | 4.201 | 2 | 0.122 |
| | Weekday | 694.182 | 12.421 | 6 | 0.053 |
| | Time Period | 699.874 | 18.113 | 4 | 0.001 |
| | Weather Conditions | 687.117 | 5.355 | 1 | 0.021 |

Figures Caption List

Figure 1. Conceptual framework for the study: overview of spatial and temporal analysis methodology.

Figure 2. Spatial distribution of crashes involving VRUs based on level of severity injury.

Figure 3. Monthly evolution of number of VRUs injured on road crashes based on level of severity injury (2012-2015)

Figure 4. Temporal distribution of vehicle-pedestrian and -cyclist crashes across different days (2012-2015).

Figure 5. Temporal distribution of vehicle-pedestrian and -cyclist crashes for different age groups (2012-2015).

Figure 6. Temporal distribution of severely injured and dead pedestrians by type of road location for Aveiro, Porto and Lisbon.

Figure 7. Temporal distribution of severely injured and dead cyclists by type of road location for Aveiro and Lisbon.

Figure 8. Relative proportion of severely injured and dead VRUs for each city under study for the build environment.

Figures

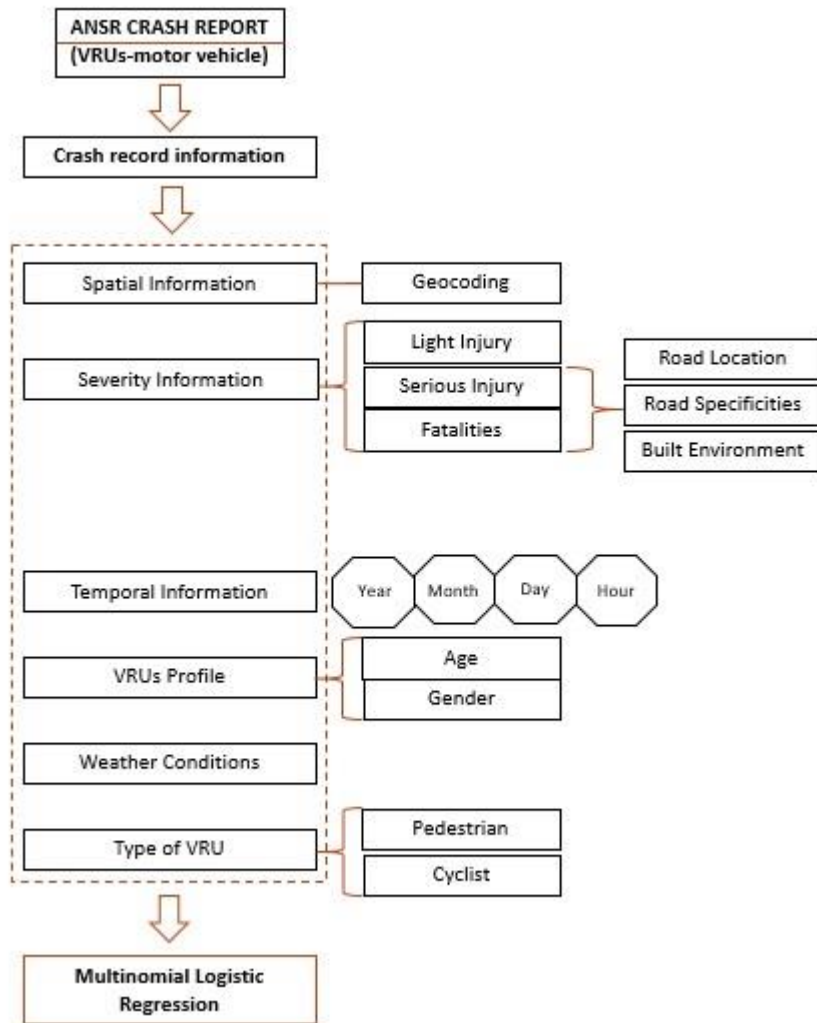


Figure 1. Conceptual framework for the study: overview of spatial and temporal analysis methodology.

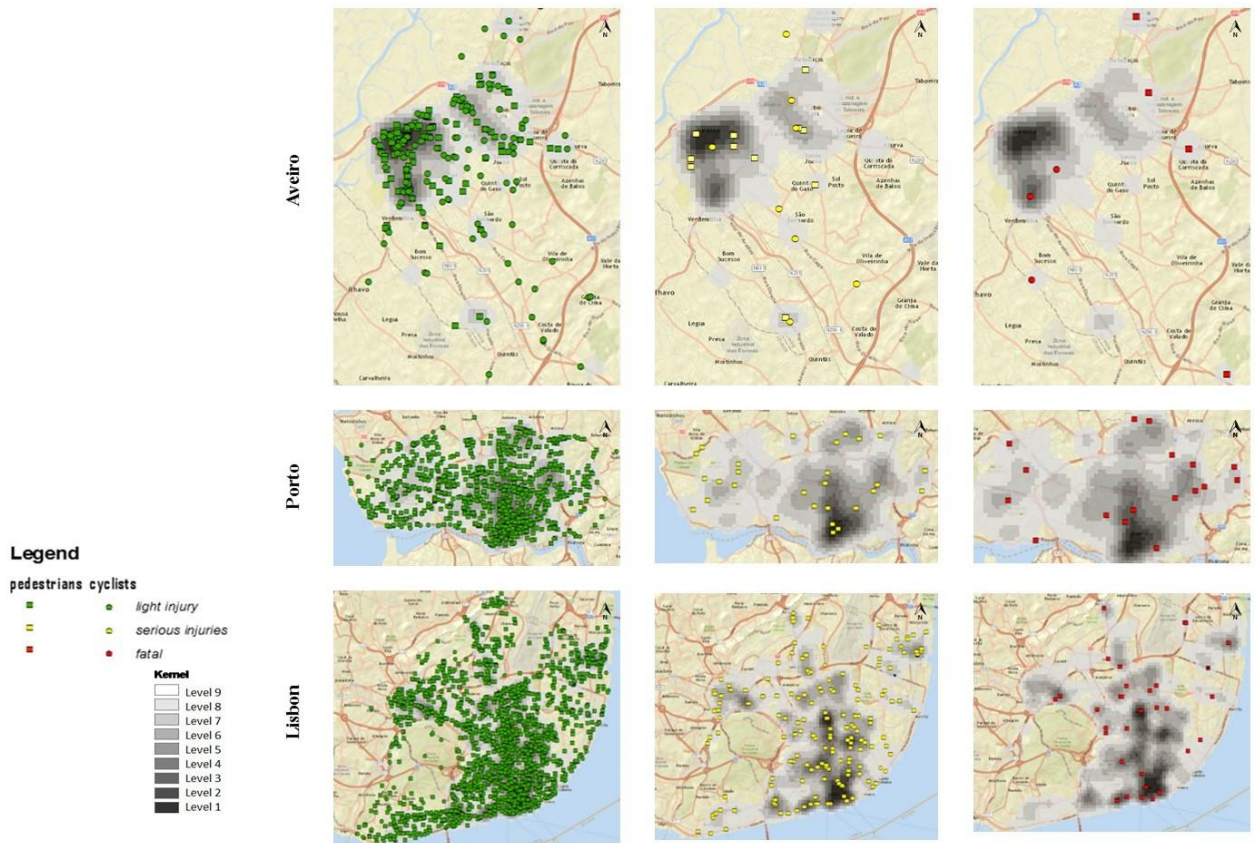


Figure 2. Spatial distribution of crashes involving VRUs based on level of severity injury

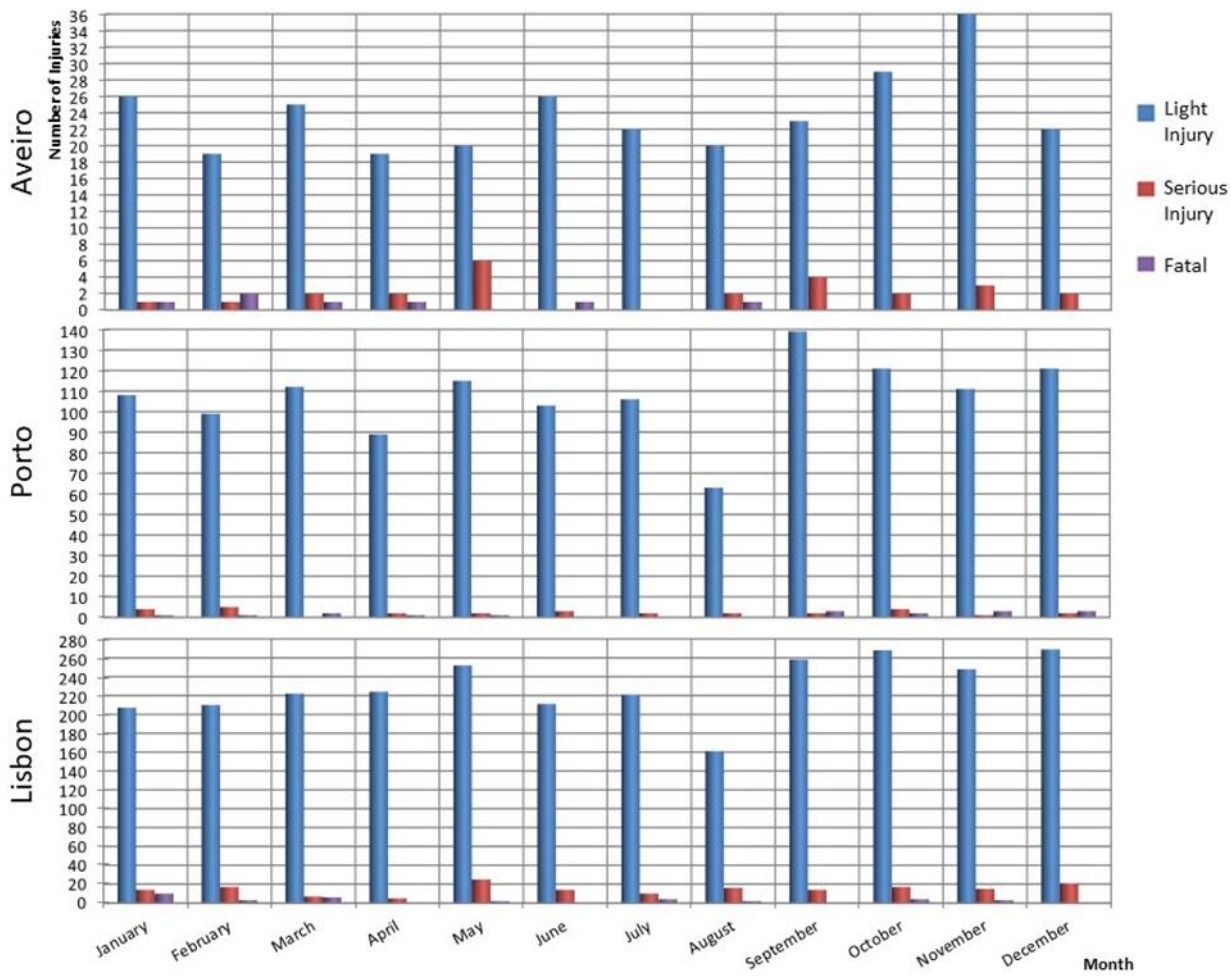


Figure 3. Monthly evolution of number of VRUs injured on road crashes based on level of severity injury (2012-2015)

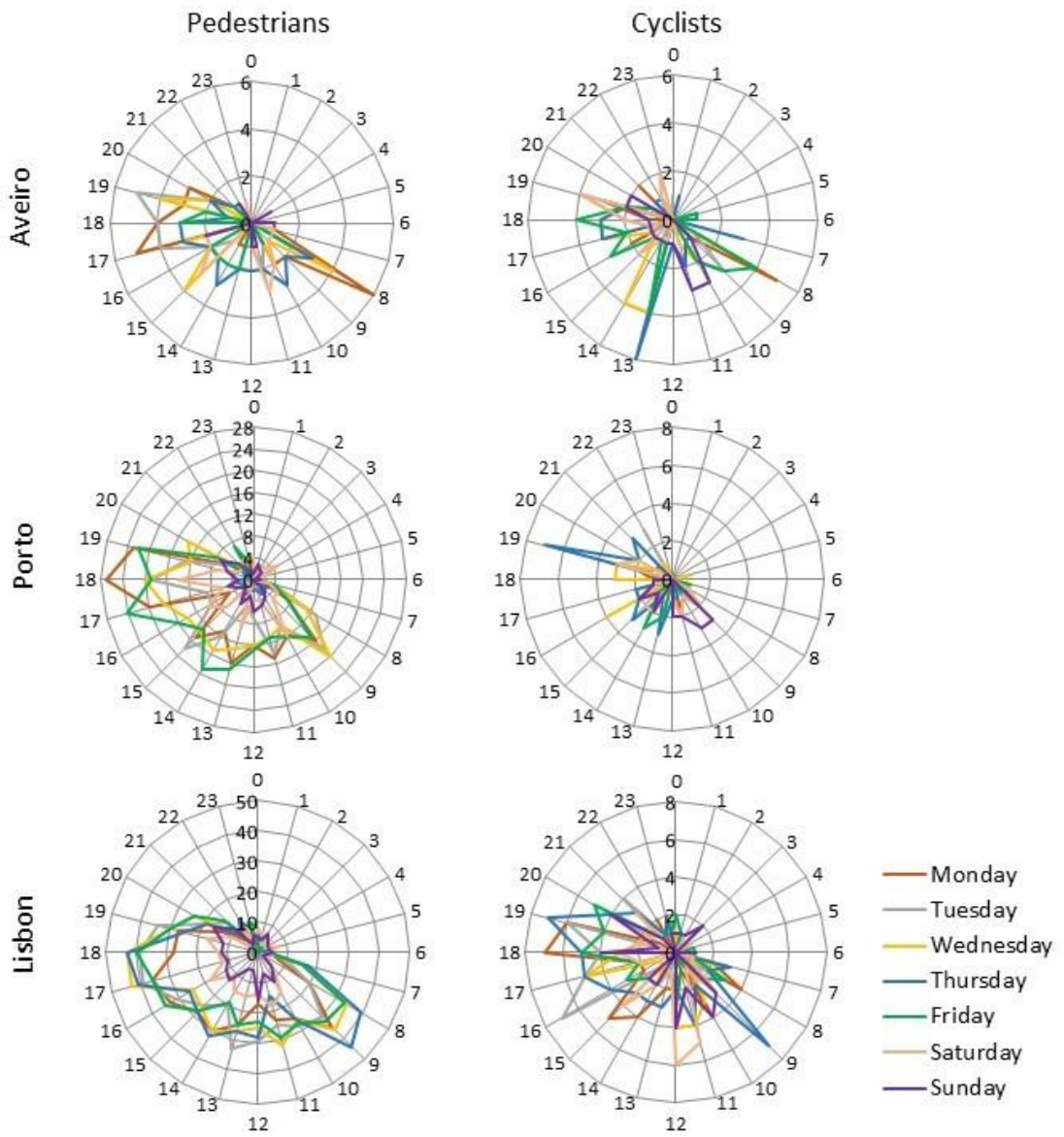


Figure 4. Temporal distribution of vehicle-pedestrian and -cyclist crashes across different days (2012-2015).

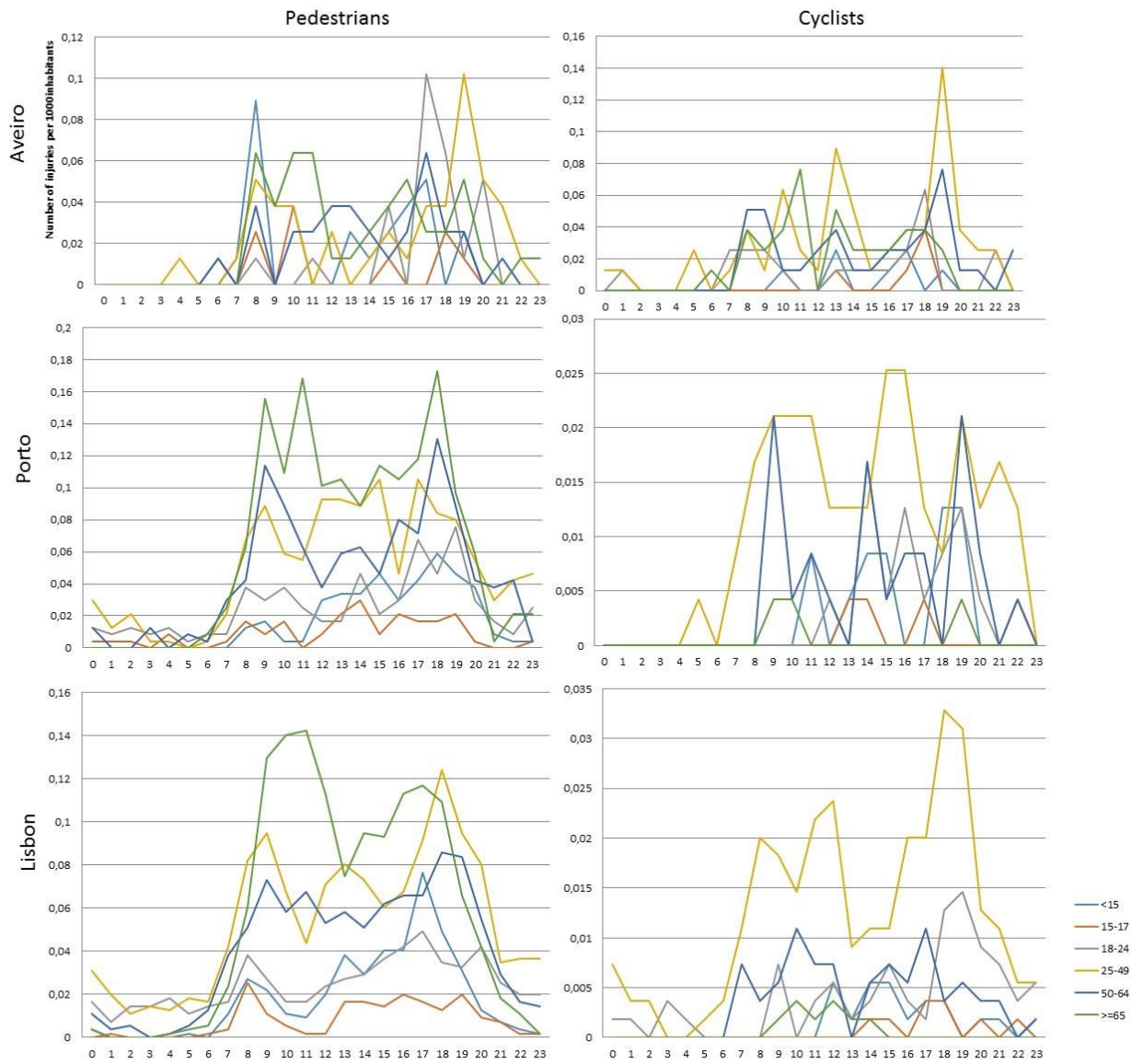


Figure 5. Temporal distribution of vehicle-pedestrian and -cyclist crashes for different age groups (2012-2015).

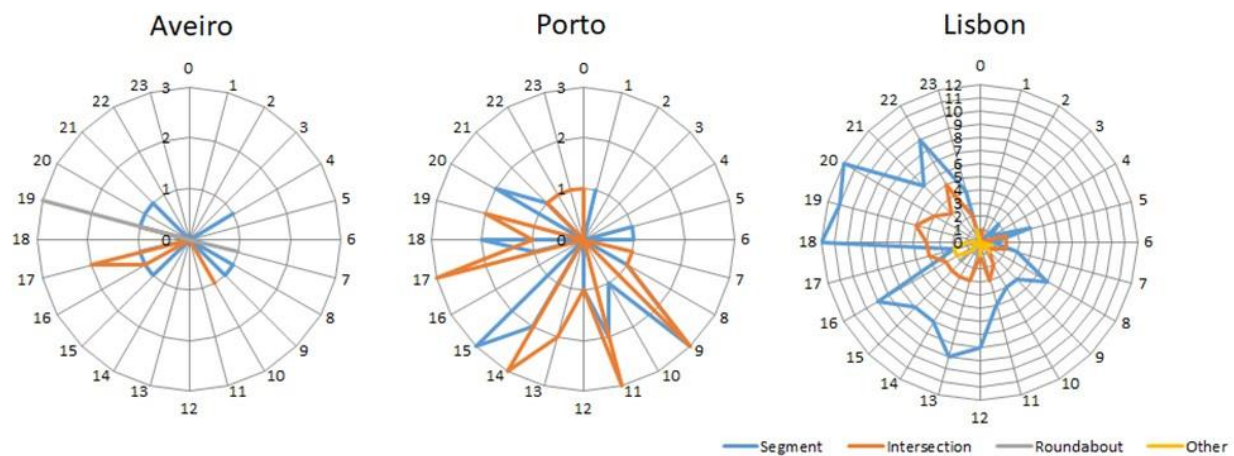


Figure 6. Temporal distribution of severely injured and dead pedestrians by type of road location for Aveiro, Porto and Lisbon.

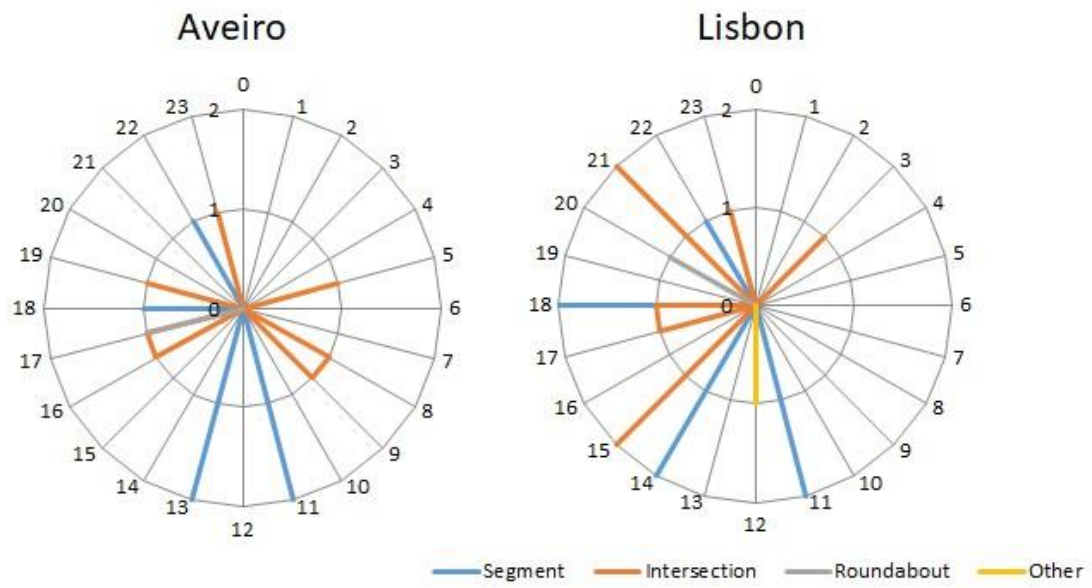


Figure 7. Temporal distribution of severely injured and dead cyclists by type of road location for Aveiro and Lisbon.

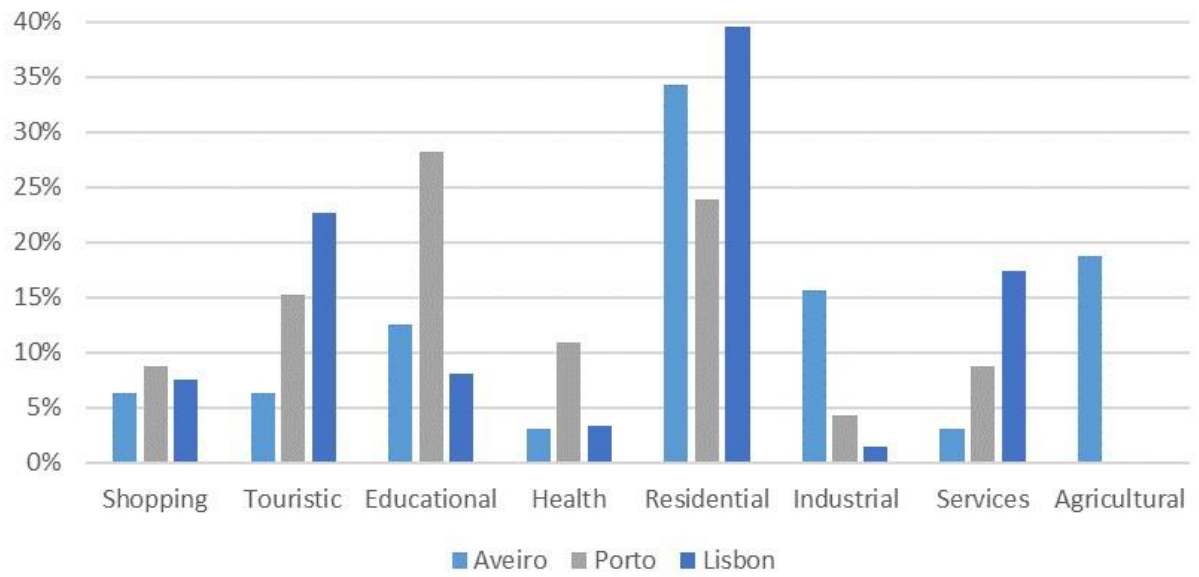


Figure 8. Relative proportion of severely injured and dead VRUs for each city under study for the build environment.