Accident Analysis and Prevention 132 (2019) 105278

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Identification of differential risk hotspots for collision and vehicle type in a directed linear network

Álvaro Briz-Redón[a,*], Francisco Martínez-Ruiz[b], Francisco Montes[a]

[a] Statistics and Operations Research, University of València, C/ Dr. Moliner, 50, 46100 Burjassot Spain
[b] Statistics Office, City Council of València

ARTICLE INFO

Keywords:
Hotspot detection
Collision type
Vehicle type
Spatial statistics
Linear network
Kernel density estimation

ABSTRACT

Traffic accidents can take place in very different ways and involve a substantially distinct number and types of vehicles. Thus, it is of interest to know which parts of a road structure present an overrepresentation of a specific type of traffic accident, specially for some typologies of collisions and vehicles that tend to trigger more severe consequences for the users being involved.

In this study, a spatial approach is followed to estimate the risk that different types of collisions and vehicles present in the central area of Valencia (Spain), considering the accidents observed in this city during the period 2014–2017. A directed spatial linear network representing the non-pedestrian road structure of the area of interest was employed to guarantee an accurate analysis of the point pattern.

A kernel density estimation technique was used to approximate the probability of risk along the network for each collision and vehicle type. A procedure based on these estimates and the sample size locally available within the network was designed and tested to determine a set of differential risk hotspots for each typology of accident considered. A Monte Carlo based simulation process was then defined to assess the statistical significance of each of the differential risk hotspots found, allowing the elaboration of rankings of importance and the possible rejection of the least significant ones.

1. Introduction

It is well known that collision and vehicle type are two capital variables that condition the severity of a traffic accident. To cite a couple of studies in this regard, Chang and Wang (2006) found vehicle type to be the most crucial factor that determines the severity of a traffic accident (which was determined to be superior for pedestrians, motorcyclists and cyclists), whereas Golob et al. (1987) verified that fixed-object and crossing were the most severe types of collisions. Thus, having the ability to detect the small sections of a road structure (which we call microzones in the remainder of the paper) that are particularly prone to present a specific type of accident is of great interest in order to implement preventive measures, specially for the type of accidents whose severity is expected to be higher. In the next paragraphs, a revision of previous studies that focused on the occurrence and causality of traffic accidents (considering the effect of collision or vehicle type) is carried out.

Regarding collision types, Ye et al. (2009) applied a simultaneous equations model to assess the frequency of different types of collisions (head-on, rear-end, sideswipe) around rural intersections, showing that a larger shoulder width associates with higher frequencies of angle, sideswipe, and pedestrian-involved accidents. Geedipally et al. (2010) studied the proportion of collisions by type with the aid of multinomial logit models. Several findings were obtained in their study: annual average daily traffic (AADT) associated to an increase of rear-ends, an increase in the percentage of trucks led to a decrease of the incidence of single-vehicle and head-on accidents (in comparison with the rear-end type) and an increase in both lane and shoulder width implied a decrease in rear-end collisions. Moreover, Geedipally and Lord (2010) proposed the distinction between single-vehicle and multi-vehicle collision in order to better model the incidence of traffic accidents given the different nature of both accident typologies. Dell'Acqua et al. (2013) defined several safety performance functions to establish the risk that associates to several collision types (head-on, rear-end, single-vehicle run-off-road) under a set of possible scenarios mainly involving surface conditions, light presence and the geometric structure of the street. Hosseinpour et al. (2014) specifically analyzed head-on collisions with a wide range of count-data models which showed that horizontal curvature, terrain type and heavy-vehicle traffic correlated with a high risk of observing this type of traffic accident. Finally, Wang et al.

* Corresponding author.
E-mail address: alvaro.briz@uv.es (Á. Briz-Redón).

https://doi.org/10.1016/j.aap.2019.105278
Received 18 January 2019; Received in revised form 3 April 2019; Accepted 19 August 2019
Available online 10 September 2019
0001-4575/

(2017) recently employed multivariate Poisson-lognormal models to analyze the factors contributing to traffic accidents with different severity, ranging from property-damage-only to fatal. In this work, the type of collision was mainly classified in terms of its directionality: same-direction, intersecting-direction, opposite-direction and single-vehicle. The results of the models dependent on these four types of collisions unveiled a differential effect of several road characteristics, as wide lanes associating with more opposite-direction accidents but less of single-vehicle type, or shoulder widths of more than 8 ft. presenting a negative association with single-vehicle collisions, among other examples.

On the other hand, many studies have focused on the investigation of a specific type of vehicle, with an special emphasis being put on the estimation of the probability of injury and the severity associated with it. For instance, some studies have investigated these effects in accidents involving bicycles (Kim et al., 2007; Walker, 2007), motorcycles (Shankar and Mannering, 1996; Savolainen and Mannering, 2007) or heavy vehicles (Chang and Mannering, 1999; Anderson and Hernandez, 2017).

In this work, a collection of traffic accidents has been projected into a road network structure, giving rise to a point pattern that lies exactly on the network. Furthermore, several informative items were available for each traffic accident in the dataset, including the type of collision, type of vehicles involved, date, time of the day and number of people that got injured. In the field of spatial statistics, such informative variables are known as the marks of the point pattern (the pattern is then called a marked point pattern). Marks represent additional information about the events that form the point pattern, which constitute random variables (before being observed) that characterize the point pattern itself (Baddeley, 2010).

The main goal of this study was the design of a procedure that allows the detection of microzones of a road network where a specific type of accident is overrepresented, focusing on two of the marks available in the dataset: collision and vehicle type. These two marks are selected given the differential risk of severity they can trigger for the people involved in a traffic accident, and the plausible influence that spatial road configurations may have in relation to certain types of collisions or vehicles.

Therefore, the paper is structured as follows. Firstly, the network structure that was used for the analysis is accurately described and located in its real context. Secondly, the traffic accident dataset that was employed is depicted, including a summary of the curation process and a detailed subsection on the information available regarding each traffic accident (by type of collision and vehicle). Then, a methodology is proposed to find microzones of the network that show a differential risk for a collision or vehicle type, including the posterior application of a Monte Carlo technique that serves to determine the statistical significance of each of these microzones. This procedure is fully displayed through an exemplification that makes use of accidents involving motorcycles. Finally, the methodology is applied and discussed for multiple collision and vehicle types.

2. Data

2.1. Network structure

A linear network composed of 1664 vertices and 2513 segments, which represented a total length of 191.14 km of road structure was analyzed. This network broadly included the city center of Valencia (Ciutat Vella District) and its five surrounding districts (l'Eixample, Extramurs, Campanar, la Saïdia and el Pla del Real), which are all shown, subdivided according to their boroughs, in Fig. 1a (OpenStreetMap contributors, 2017; Graul, 2016). This area of Valencia contains some of the most travelled avenues of the city and represents, as a whole, an homogeneous and highly connected road structure.

Network complexity was reduced without altering its basic geometrical shape with the application of a simplification algorithm that merges segments sharing a vertex of second degree (through which only two segments are connected). Moreover, network preprocessing included the slight modification of highly complex intersections, the transformation of roundabouts into simpler polygons, and the removal of pedestrian streets. The complete process was performed with the help of the R package *SpNetPrep* (Briz-Redón, 2019b).

Finally, the network was endowed with a direction according to traffic flow at this part of Valencia (also performed with *SpNetPrep*). Some of the segments of the network were defined as bidirectional, representing two-way roads where no median strip separates the two flows of vehicles. If a median strip is located in a road, two (parallel) segments are defined for the network. A representation of the final network that was employed in this study is available in Fig. 1b, including the direction of traffic flow. The few differences that can be appreciated between the administrative structure portrayed in Fig. 1a and the linear network in Fig. 1b, were only executed to give the border of the whole structure a more solid representation.

2.2. Accident dataset

A total of 11,006 traffic accidents registered by the Police Department of the city of Valencia (Spain) during the years 2014 to 2017, which took place in the roads belonging to the area of the city described in the previous section were analyzed. Each of these accidents was accurately geocoded into the network from the address information recorded by the Police officers minutes after the accident had occurred. Manual curation of the data broadly included the following steps: selection of the accidents that took place in streets located within the districts of the area of analysis, geocoding (longitude–latitude coordinates) of the accidents from the street addresses reported by the Police via the Google Maps API and the R package *ggmap* (Kahle and Wickham, 2013), revision of the obtained coordinates by applying reverse geocoding with the same package, and final inspection of the projection of these coordinates into the linear network.

As it was already mentioned in the introduction, each traffic accident in the dataset had information attached that made the point pattern become a marked point pattern. The marks regarding the collision type and the vehicles involved in each accident were chosen to be further analyzed and are described in the following two sections.

2.3. Types of collisions

Accidents are classified by Police officers according to the way they took place, leading to a quite complex variable with dozens of different categories. Considering the frequencies of these categories and examining the coincidence between some of them, only six collision types were established for the dataset: *Crossing*, *Fixed-object*, *Rear-end*, *Run-off-road*, *Run-over* and *Side*. The following lines include a brief description of each of these types.

- *Crossing*: The head of a vehicle collides to one of the sides (lateral) of another (moving) vehicle. Frontal collisions (between the front ends of two vehicles) were also assigned to this group as they were not abundant enough to establish a specific category.
- *Fixed-object*: A vehicle collides to a parked vehicle or to any fixed element in the street.
- *Rear-end*: The head of a vehicle collides to the end of another (moving) vehicle.
- *Run-off-road*: A vehicle loses control and leaves its traffic way, possibly invading the opposite direction of traffic flow or the sidewalk.
- *Run-over*: A vehicle hits a person and possibly drives over him/her. Sometimes it is also employed when a vehicle hits another one of clearly lower dimension and/or weight.
- *Side*: Lateral collision between two or more vehicles.

Fig. 1. Boroughs of the area of the city of Valencia that was studied (a) and complete network structure used for the analysis, with arrows indicating the traffic flow (b). In (a), the nomenclature of each borough is overlayed on the map, with the first number indicating the district the borough belongs to, from 1 to 6: Ciutat Vella, l'Eixample, Extramurs, Campanar, la Saïdia and el Pla del Real. (For the visualization of (b), the reader is referred to the web version of this article.)

Table 1
Sample sizes, relative frequencies and Moran's I values (with associated p-values testing the null hypothesis of no spatial autocorrelation) of all the categories available for the two marks considered.

Mark	Category	n	%	I (p-value)
Collision type	Crossing	3082	28.00	0.04 (0.00)
	Fixed-object	1518	13.79	0.06 (0.00)
	Rear-end	2691	24.45	0.16 (0.00)
	Run-off-road	198	1.80	0.03 (0.05)
	Run-over	701	6.37	0.13 (0.00)
	Side	1314	11.94	0.08 (0.00)
Vehicle type	Bicycle	635	5.77	0.09 (0.00)
	Car	9524	86.53	0.11 (0.00)
	Lorry	420	3.82	0.03 (0.05)
	Motorcycle	2811	25.54	0.12 (0.00)
	Private bus	212	1.93	0.11 (0.00)
	Public bus	747	6.79	0.21 (0.00)
	Van	916	8.32	0.05 (0.00)

It must be remarked that some traffic accidents can be quite complex and involve several types of collisions and situations, making the treatment of this type of data quite challenging. In fact, a part of the accidents that took place in the road network considered during the period 2014–2017 remained without collision type being assigned due to the lack of information available in this regard, or because the literal concept annotated was completely different to the six predominant types that have been just defined (avoiding a proper recodification of the collision type). Table 1 shows the relative frequencies of the different collision types in the dataset (they represent 86.35% of the total because of the accidents uninformed or categorized with an infrequent type of collision) and the Moran's I values (Moran, 1950a,b) shown by each of the patterns, which confirm the significant spatial autocorrelation of every type of collision (at the 0.05 level, even though the *Run-off-road* type is at the limit).

2.4. Types of vehicles

Information regarding the types of vehicles implicated in each accident was also available. As expected, cars were the vehicles that appeared more frequently in the dataset by far, followed by motorcycles (mopeds were also recoded to this type). The involvement of cars in most of the traffic accidents led to discard its specific study, putting the focus on the rest of vehicle types.

Particularly, the other types of vehicles present in the dataset were bicycles, buses (private and public), lorries and vans. Accidents implicating public buses were separated from those involving private ones because the former are subject to specific routes, and also because of their singular importance as part of the public transportation system of the city.

It is needed to say that this mark of the point pattern was used in the form of a set of binary variables referring to each of the types of vehicles registered in the dataset. Therefore, each accident allowed the presence of several vehicle types, a situation that did not hold for the collision type. As a consequence, the addition of the relative frequencies in Table 1 exceeds 100.

The possibility that some microzones of the road network are particularly dangerous for some of these vehicle types was deeply analyzed, which seems reasonable according to the high values of the Moran's I in Table 1 that indicate the presence of spatial aggregation, specially for public buses.

3. Methodology

3.1. Software

The R programming language (3.4.1 version, R Development Core Team, Vienna, Austria) (Core Team, 2017) was used to obtain all the results presented in this work. The R packages *ggmap* (Kahle and Wickham, 2013), *spatstat* (Baddeley et al., 2015), *spded* (Bivand and Piras, 2015) and *SpNetPrep* (Briz-Redón, 2019b) were specifically required for some parts of the analysis.

3.2. Kernel density estimation

Kernel density estimation (KDE) is typically used in spatial statistics to estimate the intensity of a point pattern over a geographical space. Particularly, KDE is also applied to estimate the intensity of a pattern along a linear network, a process that requires some modifications of the classical formulas to accommodate the particularities of this spatial structure (Okabe et al., 2009; Okabe and Sugihara, 2012). In this regard, McSwiggan et al. (2017) recently proposed an equal-split continuous version of a kernel density estimator for linear networks, which was the one considered for this study. This estimator can be computed as follows for any point x of the network:

$$\lambda_\sigma(x) = \sum_{z \in A(x,\sigma)} k(d_L(x,z)) a^C(\pi) \quad (1)$$

where σ is the kernel's bandwidth, $A(x, \sigma)$ is the set of points of the network where an accident took place that were situated at a distance from x lower than $k(w) = 1/\sigma\sqrt{\pi}e^{(-w/\sigma)^2}$ is the kernel function (Gaussian), $d_L(x, z)$ represents the shortest-path distance along the network from x to z and $a^C(\pi) = \prod_{j=1}^{m} 2/(\deg(v_j))$, where $\pi = [v_1, \ldots, v_m]$ denotes the set of vertices of the network that have to be passed through to travel the shortest path that joins z with x and deg() represents the

degree of a vertex of the network, meaning the number of segments incident to the vertex. It is of need to remark that the computation of the distance d_L between any two points of the network takes network's directionality into account, providing a realistic measure of the distance between the two points according to traffic flow.

The Gaussian kernel was used for simplicity, due to the known fact that this choice has little effect on the results (Silverman, 2018). Bandwidth parameters ranging from 50 m to 150 m were tested in order to check the effect of this election.

3.3. Estimating a relative probability for each type of accident

KDE can be used to estimate the spatially-varying relative probability of occurrence for a type of event or the relative risk between several event types, both two applications of special interest in case-control and related studies (Kelsall and Diggle, 1998; Diggle et al., 2005; Serra et al., 2013). Hence, if $\{y_i\}_{i=1}^n$ represents the binary outcomes (for example, case or control) of a collection of events observed at points $\{x_i\}_{i=1}^n$, KDE allows the derivation of a risk surface (Kelsall et al., 1995; Kelsall and Diggle, 1995) over space that can be interpreted as a conditional probability of observing a case ($Y_i = 1$) at a location X_i, with X_i and Y_i representing the random location of a spatial event and its outcome, respectively. In addition, in a more general setting, the case-control situation can be extended and one can set $y_i = 1$ if one mark of the point pattern takes certain value at point x_i, and 0 otherwise. This approach is taken in our paper, which enables to define a risk surface regarding the involvement of a specific collision or vehicle type in the neighbourhood of a point x_i.

Therefore, according to the formulas derived by Copas (1983) and Kelsall and Diggle (1998) for planar point patterns, an estimate (that depends on the bandwidth σ) of the risk presented by a type of event at a location x is:

$$p_\sigma(x) = \frac{\sum_{i=1}^n K_\sigma(x - x_i) y_i}{\sum_{i=1}^n K_\sigma(x - x_i)}$$

where n is the number of events observed and $K_\sigma(u) = h^{-2}K(\sigma^{-1}u)$ is a kernel function with $K(u) = (2\pi)^{-1}\exp(-1/2||u||^2)$ (Gaussian), being $||\cdot||$ the Euclidean norm. This formula is then adapted to the case of the network KDE, which leads to the estimation of a relative probability of risk for any typology of accident at a location x of the road network:

$$p_\sigma(x) = \sum_{i=1}^n \lambda_\sigma(x_i) y_i \Big/ \sum_{i=1}^n \lambda_\sigma(x_i) \quad (2)$$

where $\lambda_\sigma(x)$ follows Eq. (1). Hence, $p_\sigma(x)$ approximates the relative probability for the typology of traffic accident being represented by $y_i = 1$ to be observed at location x, which relies on the information provided by all the traffic accidents occurred within a linear radius (following the network structure) of σ meters from x.

3.4. Detecting differential risk hotspots

The main objective of the study was to design a methodology capable of identifying microzones of a road network where certain accident typology is overrepresented. From now on, these microzones will be referred to as differential risk hotspots (see Fig. 2 for a full graphical description). Many previous studies have already dealt with the accurate detection of hotspots at the road segment level (Xie and Yan, 2013; Nie et al., 2015; Harirforoush and Bellalite, 2016), but focusing on a type of accident and assessing risk in relation to other types has been less investigated by far. Therefore, the following paragraphs include a description of the procedure proposed for differential risk hotspot detection, which relies on the estimation (across space) of a relative probability of occurrence for each event type, according to Eq. (2). An implementation of this methodology is available in the R package DRHotNet, which has been recently released (Briz-Redón, 2019a).

The first step of the procedure consists in estimating (with Eq. (2)) the relative probability of a specific accident typology along the complete spatial network. Such estimates need to be computed at a partition of the whole road network in order to obtain a estimation of the risk that the type of accident presents across space. In particular, the middle points of the 2513 segments of the road network from Valencia could have been chosen for this step, but in order to gain accuracy, specially near and around road intersections, the linear network was subdivided into shorter segments that are called *lixels* in literature (Xie and Yan, 2008) (Fig. 2b). A value of 50 m was chosen such that the length of the segments of the new network (from now referred to as the *lixellized* network) did not exceed this threshold. The resulting lixellized network presented a total of 5099 segments and 4250 vertices, in which all traffic accidents available of any type were projected (Fig. 2c).

Once the probabilities of risk were computed for a typology of accident at the middle points of the lixellized network (Fig. 2d), several approaches were tested for attempting the definition of representative hazardous hotspots of road segments for each accident's typology. Firstly, following one of the most usual methodologies, local Moran's I (LISA) statistic values (Anselin, 1995) were computed for each of the lixels. Then, lixels showing a significant local association at the 0.05 level were selected and grouped according to their contiguity, leading to potential differential risk hotspots. However, probably due to the only moderate spatial autocorrelation showed by some of the types of accidents along the network and also to the small sample size at many microzones, this approach did not provide satisfactory results.

Therefore, an alternative method was finally established, which consisted in selecting the road segments with a superior relative probability estimate in comparison with the mean value for the whole network. More specifically, road segments associated with a relative probability exceeding the mean value for the network in more than k times the standard deviation presented by all the probabilities estimated were preselected. From this set of pieces of the network of length up to 50 m, only those where n or more accidents had occurred within a linear radius of σ meters (the bandwidth value chosen for estimating the probabilities) were finally selected for the construction of the hotspots (Fig. 2e). Thus, the inclusion of the n parameter allows discarding some estimated probabilities that are not based on a large enough dataset of traffic accidents, which could be artificially large and consequently meaningless.

A sensitivity analysis was carried out on k and n with several values of the bandwidth parameter in order to find sensible choices of these two parameters that allow the obtention of a reasonable set of differential risk hotspots. The prediction accuracy index (PAI) developed by Chainey et al. (2008), which has already been used in some studies involving traffic accidents (Thakali et al., 2015) or crimes (Van Patten et al., 2009), was slightly modified to perform this analysis. Hence, a type-specific PAI (PAI_{type}) was defined as follows:

$$\text{PAI}_{\text{type}} = \frac{n_{\text{type}}/N_{\text{type}}}{m/M}$$

where n_{type} is the number of traffic accidents recorded in the hotspots with the type of interest, N_{type} the total number of traffic accidents in the study with the type of interest, m is the length (in meters) of all the hotspots detected, and M is the total length of the network structure being considered (191.14 km).

Thus, the PAI_{type} index computes the ratio between the proportion of accidents that took place in the set of differential risk hotspots obtained and the proportion of network length that is spanned by these hotspots. A higher value of the PAI_{type} index indicates that the hotspots found have proportionally condensed more traffic accidents for a given length of road, which represents a better performance in this context.

3.5. Assessing hotspot significance

The use of a KDE technique facilitates the estimation of the intensity

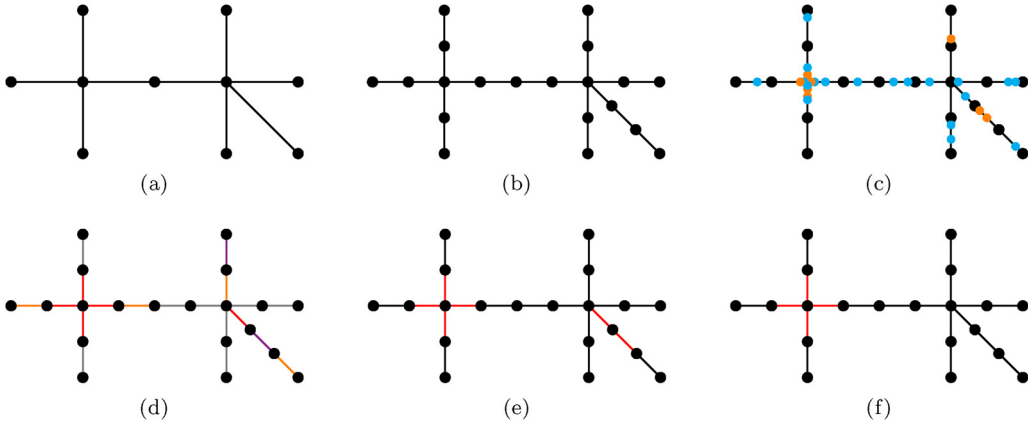


Fig. 2. Graphical description of the procedure implemented in order to detect differential risk hostpots. Starting from a road network of interest (a), this is segmented into shorter elements called lixels (b). The traffic accidents available are then projected into the lixellized network in (b), some of which can be of one specific type (orange points), whereas the rest are of any other type (blue points), as shown in (c). Next, KDE is employed to produce an estimation of the probability of risk for the accident type being studied (symbolized by the orange points) at each of the middle points of the lixellized network (d). As an illustration, in (d) the lixels are coloured according to a sensible guess of these probabilities in view of the point pattern in (c), ranging from average or below average (gray) to very high (purple). The lixels satisfying the conditions imposed by the k and n parameters are selected (in red) and grouped, becoming differential risk hotspots (e). Finally, a Monte Carlo technique is applied to the hostspots in (e), yielding a value of significance for each of them that can occasionally lead to their rejection (f). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of a point pattern over a space and the detection of microzones that show high/low values of intensity in comparison with their surroundings. However, the mere use of KDE does not provide any statistical significance value that helps to discriminate which hotspots show the most remarkable differential behaviour for a type of collision or vehicle. In this regard, the methodology introduced by Bíl et al. (2013) is partially imitated to find which of the predefined differential risk hotspots can be truly considered as a microzone of high dangerousness for a specific typology of accident, and not a consequence of the small sample size that the pattern presents at some microzones of the network (even though a right choice of the n parameter reduces the chances of this undesired possibility). Hence, whereas the procedure described in the previous section serves to obtain a set of differential risk hotspots, the one described in the present section enables to assign a statistical significance value to each of these hotspots. Then, coldspots can be ranked according to their importance, or even be rejected if they do not show enough statistical evidence of presenting a differential risk for the type of traffic accident being studied.

The methodology that is defined to check the statistical significance of each differential risk hotspot is now described. First, the point pattern available is left fixed in space whereas the marks (collision and vehicle type) of the events are changed randomly a total of 750 times. Following a Monte Carlo approach in order to test the null hypothesis of random mark assignment (meaning random type of collision or vehicles implicated), for every single simulation, and for every differential risk hotspot being analyzed, the probabilities of risk are computed at the middle points of the lixels forming each hotspot (considering each simulated marked point pattern). Then, a probability of risk for the type of accident is assigned to each hotspot at every simulation by simply averaging the probabilities estimated for every lixel (the average is weighted by lixel's length to increase the contribution of longer lixels, even though this decision barely alters the results). This process generates an empirical distribution for the probability of risk at every differential risk hotspot, allowing the construction of a statistical significance measure (a p-value) for each of them. The assessment of a statistical significance to each differential risk hotspot can eventually lead to discard some of them or, at least, reduce their importance in favor of the most significant ones (Fig. 2f).

On the choice of the value 750 as the number of simulations performed to assess statistical significance, the study of Robey and Barcikowski (1992) was followed. Assuming a type I error $\alpha = 0.05$, a type II error $\beta = 0.2$ (which implies a power for the test of 0.8) and a "liberal criterion" according to the definition of robustness for a statistical test introduced by Bradley (1978), a value around 750 would be recommended. A larger choice may be better depending on the level of uncertainty one wishes to undertake, but here the definition of 750 iterations was considered fair enough according to the results that were obtained and the computation time that was saved avoiding the use of a more stringent criterion.

3.6. Ranking hotspots according to further criteria

The detection of microzones of a road network presenting a singular risk for a collision or vehicle type should be followed by the implementation of a set of countermeasures that attempt to improve their safety. As stated, the procedure introduced in this paper allows the ranking of the differential risk hotspots found according to their empirically determined statistical significance.

However, it would be sensible to bring into the equation several external factors, not considered explicitly by our procedure, that are capital to establish a systematic decision-making process that optimizes the social cost-benefit associated to the application of traffic safety countermeasures. Hence, the absolute number of traffic accidents found in a hotspot, the proportion of severe accidents observed or some collision/vehicle types particularly overrepresented (which are highly correlated with severity outcomes) are only a few of the factors that may be incorporated for this matter. In this regard, some methodologies for hotspots ranking that allow the combination of several indicators, like the one developed by Coll et al. (2013), would be useful.

4. Results and discussion

This section starts with a subsection that illustrates the methodology proposed in this paper by showing all the steps that were followed to detect microzones of the linear network where traffic accidents involving motorcycles were overrepresented. Furthermore, this section includes a discussion on the choice of the k and n parameters and their appropriateness depending on the necessities of the researchers and professionals involved in the design of preventive measures. The second subsection contains a summary and analysis of the differential risk hotspots that are obtained for all the typologies of traffic accidents considered in the study.

4.1. Example of application: motorcycle accidents

Motorcycles contributed to 2811 of the 11,006 accidents available in the dataset, representing 25.54% of the total of accidents. Following Eq. (2), the relative probability of observing an accident involving a

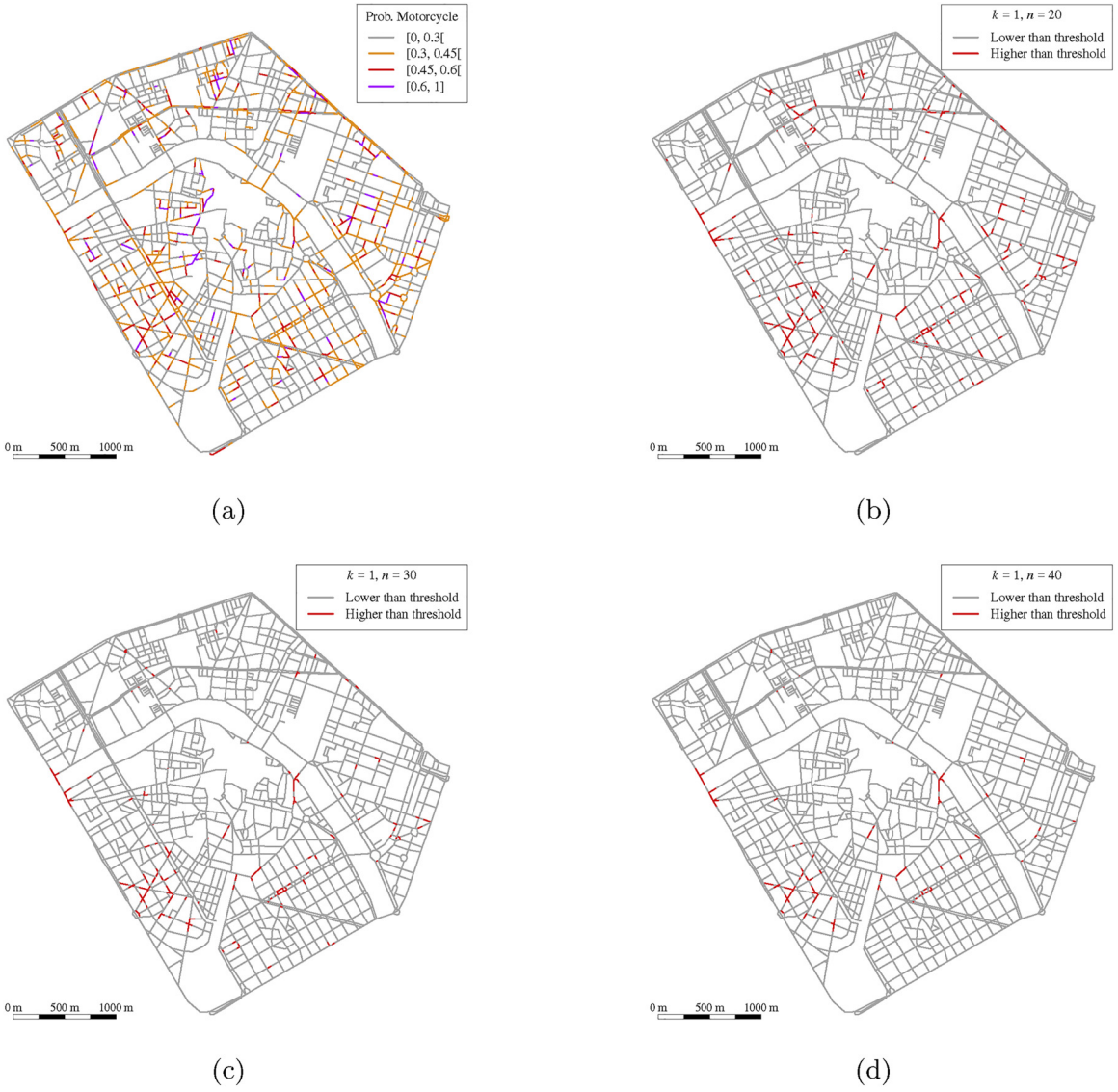


Fig. 3. Estimated probabilities of observing an accident that involves a motorcycle for each lixel of the linear network (a) and differential risk hotspots detected after the application of the methodology with $k = 1$ and (b) $n = 20$, (c) $n = 30$ and (d) $n = 40$.

motorcycle was estimated along the complete lixellized network with a bandwidth value of $\sigma = 100$ m (Fig. 3a displays a graphical representation of these estimations, with the lixels presenting a value higher than 0.3 being coloured accordingly). Several bandwidth parameters were tested, ranging from 50 to 150 m, and a value of 100 m was found a suitable choice according to the results obtained. The use of bandwidths lower than 100 m, specially those that are close to 50 m, leads to the frequent generation of many short hotspots that fail to connect to other microzones of the network (reducing the comprehensibility of the results), whereas the election of a larger value around 150 m seems to excessively extend the effect of the point pattern, likely producing fictitious microzones of differential risk.

As an illustration, the application of the hotspot detection methodology with $k = 1$ and several values of n is shown in Fig. 3b–d. As expected, for a fixed k, a larger value of n is more restrictive and less microzones of the network are pointed out.

The effect of the election of k and n is further investigated with the use of a set of values in the range [0, 2] for k and varying n from 10 to 50. Fig. 4 shows the values of PAI_{type} (Fig. 4a) and the proportion of accidents involving a motorcycle (Fig. 4b) for the sets of hotspots determined for the different values of k and n. Other vehicles (and types of collisions) were investigated with such graphical descriptions and provided similar results, so the arguments stated in the next paragraphs that are based on the case of motorcycle accidents stay true in general.

As it has already been explained, PAI_{type} index measures the ratio between the proportion of accidents of interest found within a set of hotspots (in this example, presenting a differential risk for motorcycles) and the proportion of length that these hotspots represent in the whole network structure. Therefore, the highest value of PAI_{type} would indicate the k and n values that optimize the procedure, in the sense of providing a minimal road length structure (the set of hotspots) given the number of motorcycle accidents that occurred on it. Fig. 4a suggests that a choice around $k = 1.5$ and $n = 45$ would be the optimal for PAI_{type} in this example. Moreover, in this study the interest lies in the relative probability of accident for a specific type of collision or vehicle, more than on the microzones of the network that contain the majority of them (PAI_{type} focuses on this). In this regard, Fig. 4b shows how the proportion of accidents involving a motorcycle increases as k and n do, although a very restrictive selection of k and n would imply a too reduced number of hotspots.

In conclusion, the variation of the k and n parameters can lead to very different sets of differential risk hotspots. A value for k in the

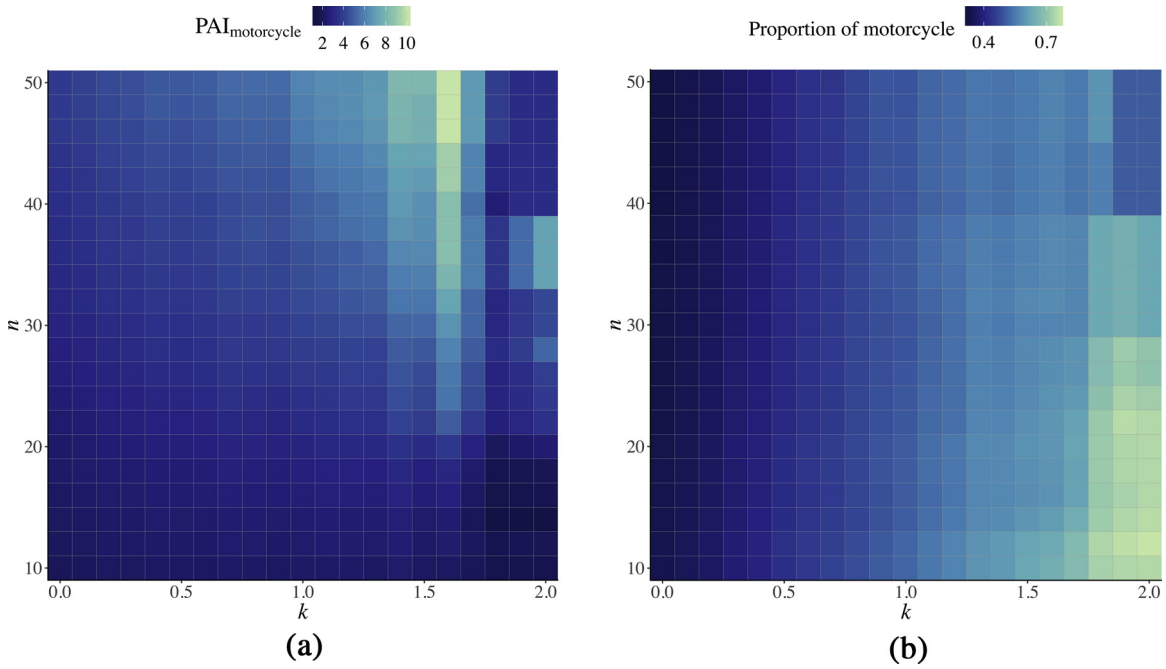


Fig. 4. $PAI_{motorcycle}$ results (a) and proportion of traffic accidents involving a motorcycle (b) within the set of hotspots obtained with different values of k and n and a bandwidth value of $\sigma = 100$ m.

Table 2
Description of the differential risk hotspots (denoted simply as hotspots in this table) that are detected when applying the procedure with different values of k and n for $\sigma = 100$ m, where $N_{motorcycle}$ is the number of accidents involving motorcycles in the complete set of hotspots for two given values of the parameters, N is the total number of traffic accidents recorded in the same space, and $p_{motorcycle}$ is the proportion $N_{motorcycle}/N$. It can be appreciated that for some combinations of these two parameters no hotspots are determined.

k	n	Hotspots	$N_{motorcycle}$	N	$p_{motorcycle}$	Length (m)	$PAI_{motorcycle}$
0.5	10	287	1247	2936	0.42	40,692.88	2.08
1	10	181	583	1128	0.52	18,217.86	2.18
1.5	10	98	200	320	0.62	6559.52	2.07
2	10	41	55	74	0.74	2320.5	1.61
2.5	10	14	14	15	0.93	588.62	1.62
0.5	20	235	1027	2461	0.42	26,287.38	2.66
1	20	124	443	881	0.50	10,277.83	2.93
1.5	20	48	129	220	0.59	2754.69	3.18
2	20	14	20	27	0.74	555.54	2.45
2.5	20	4	5	5	1.00	125.80	2.70
0.5	30	183	862	2107	0.41	17,272.68	3.39
1	30	85	347	708	0.49	5853.93	4.03
1.5	30	25	100	178	0.56	1308.92	5.19
2	30	5	10	16	0.62	159.03	4.28
2.5	30	0	–	–	–	–	–
0.5	40	128	704	1739	0.40	12,032.39	3.98
1	40	50	277	568	0.49	3824.72	4.92
1.5	40	16	84	155	0.54	880.23	6.49
2	40	2	4	8	0.50	86.04	3.16
2.5	40	0	–	–	–	–	–
0.5	50	101	583	1434	0.41	8262.58	4.80
1	50	35	240	489	0.49	2688.96	6.07
1.5	50	13	84	154	0.55	701.51	8.14
2	50	2	4	8	0.50	86.04	3.16
2.5	50	0	–	–	–	–	–

interval [1,2] and $n \geq 30$ seem appropriate to obtain a reasonable number of microzones along the network presenting differential risk for one specific type of collision or vehicle, but the final election should remain to the decision of the researchers, and specially the professionals that could be in charge of establishing preventive measures according to the results obtained (police officers and traffic experts). To illustrate this idea, Table 2 describes the results that were obtained from the differential hotspot detection procedure for several values of k and n (still with accidents involving motorcycles). With the exception of the most restrictive parameter values, the microzone that is spanned by the set of hotspots extends for some (or many) kilometers, suggesting the use of a more exigent combination if the wish is to focus only in the most relevant parts of the network for the accident typology of interest.

Keeping this in mind, the procedure here described for motorcycle accidents was performed for most of the possible outcomes of the two marks attached to the point pattern (collision and vehicle type) with $k = 1$, $n = 40$ and $\sigma = 100$, including the posterior application of the Monte Carlo technique to obtain information regarding significance. These results are shown in the next section.

4.2. General application of the methodology

The procedure is employed with the different types of vehicles and collisions informed at the available dataset, providing the significant differential risk hotspots that are shown in Figs. 5 and 6 (at the 0.05 level). As in the example included in the previous section, the values $k = 1$ and $n = 40$ were chosen in order to obtain a initial set of differential risk hotspots. Later, the statistical significance of these hotspots was assessed by the Monte Carlo technique, yielding the final collection that appears in Figs. 5 and 6 (only those with a p-value lower than 0.05 are selected). Specifically, in the case of motorcycle accidents, Fig. 6a includes the hotspots from the group shown in Fig. 3c that were yielded significant by the Monte Carlo procedure, even though the ones rejected could also be reconsidered once the former are inspected and treated by the authorities.

Occasionally, the differential risk procedure yields differential risk hotspots formed by only one lixel, which are specially notorious in the less frequent types of accidents that naturally involve a minimal number of hotspots. The examination of some of this one-lixel hotspots unveils some problematic situation that needs to be explained. The use of the KDE technique with a bandwidth of 100 m implies that the probability of risk that is estimated at the middle point of each lixel of the network is based on a slightly wider microzone that the lixel itself

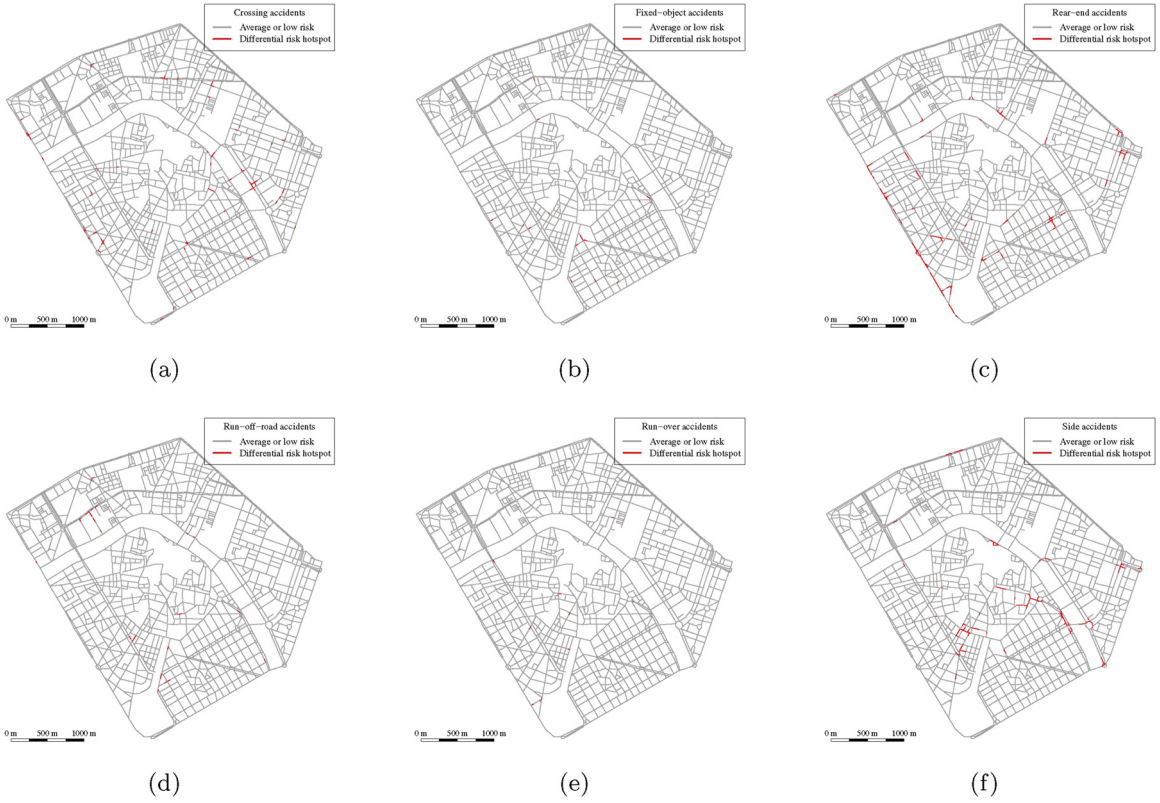


Fig. 5. Differential risk hotspots that are statistically significant (at the 0.05 level) for each collision type after the application of the detection procedure with $k = 1$, $n = 40$ and $\sigma = 100$.

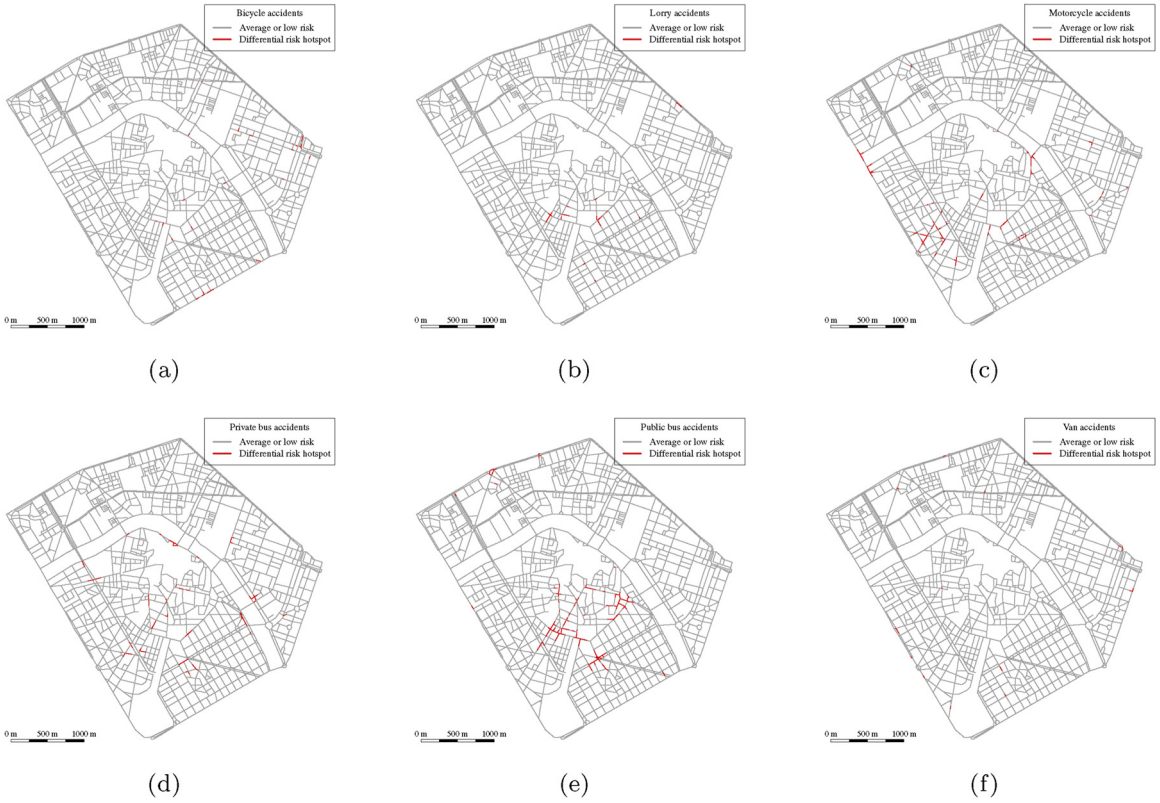


Fig. 6. Differential risk hotspots that are statistically significant (at the 0.05 level) for each vehicle type after the application of the detection procedure with $k = 1$, $n = 40$ and $\sigma = 100$.

(which cannot exceed a length of 50 m by construction). This situation can eventually lead to the declaration of a singular lixel of the network as a differential risk hotspot, although only a little number of accidents of the type being studied have occurred within that lixel. Even though such a hotspot could be simply a false positive produced by the methodology being proposed, the problem addresses automatically sometimes if one thinks about the middle point of the lixel as the center of a linear radius of 100 m along the network (following traffic flow) where that type of accident is overrepresented. For instance, this situation takes place with the one-lixel differential hotspot detected for bicycle accidents in Borough 2.2 and with the one found in the border of the Boroughs 3.1 and 3.3 related to run-over collisions (see Fig. 1a for locating these boroughs). In the former, no accidents involving bicycles were observed from 2014 to 2017, but 6 out of the 42 accidents that occurred in a radius of 100 m from the middle point of this lixel in the same period of time implicated a cyclist. Indeed, the estimated relative probability for an accident in this lixel to involve a bicycle was 0.17 (very close to the simple proportion), which is far superior to the proportion of bicycle accidents in the whole dataset (0.06). Similarly, no run-overs were recorded in the aforementioned differential hotspot regarding this kind of collision, but in a radius of 100 m from its middle point a total of 11 out of 56 traffic accidents belonged to this type (the estimation of the relative probability of a run-over collision at this lixel was 0.21, much larger than the proportion of 0.06 for the complete network).

Therefore, one should consider each of the differential hotspots pointed out by the procedure as the core of a slightly extended microzone of the network that manifests a differential risk for one specific type of accident. This way of thinking about the hotspots is usually unnecessary, as the lixels that form the hotspot normally encompass most of the accidents recorded in that microzone of the network, but becomes a requirement to give sense to the shortest hotspots that are composed by only one lixel. In the same vein, with regard to the interpretation of a particular differential risk hotspot, it is worth to remark that the type of collision should be also accounted. For instance, for run-off-road accidents, a detected microzone may be the consequence of a triggering condition located dozens of meters apart, in the same or in a connected road, which is where countermeasures should really focus.

As a summary, Table 3 describes the differential risk hotspots that were found for the different collision and vehicle types. This table includes the number of traffic accidents of each type that were recorded within the corresponding set of hotspots and the proportion of accidents they represented along this microzone of the road network. In order to minimize the possible presence of differential risk hotspots that point out a very specific microzone of the network but are really representing a wider part of it (issue treated in the previous paragraphs), the number of accidents of each type occurred within an extension of the hotspots is also indicated (a 75 m linear radius from the extreme points that include the complete hotspot).

Table 4
Description of the sets of differential risk hotspots found for each collision and vehicle type considering the percentage of traffic accidents (for the corresponding type) and road length (in relation to the whole network) that the set represents. The quotient of these two percentages is equivalent to the PAI_{type} value associated.

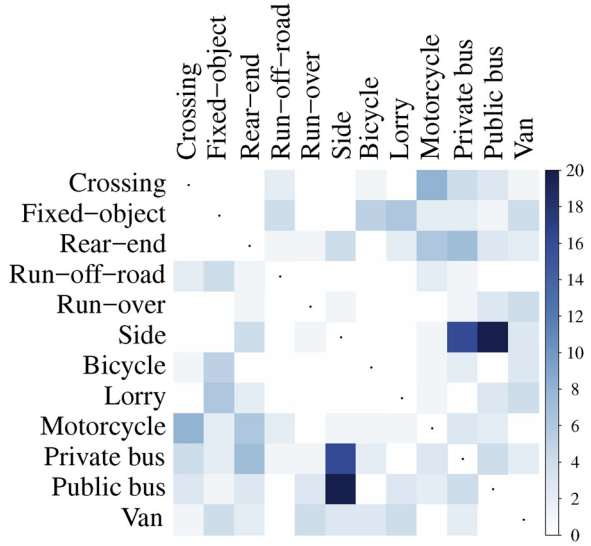
Type	% Traffic accidents	% Road length	PAI_{type}
Crossing	8.14	1.55	5.25
Fixed-object	1.78	0.33	5.37
Rear-end	15.20	2.69	5.64
Run-off-road	7.58	0.59	12.83
Run-over	1.57	0.20	7.71
Side	18.04	2.31	7.79
Bicycle	7.09	0.56	12.72
Lorry	8.10	0.51	15.79
Motorcycle	9.36	1.73	5.41
Private Bus	21.23	1.57	13.49
Public Bus	30.92	2.82	10.96
Van	3.38	0.40	8.43

Regarding the PAI_{type} values that were obtained, Table 4 indicates that these ranged from 5 to 16, approximately, being specially high for run-off-road collisions, bicycles, lorries and buses (private and public). Hence, the detection of differential risk hotspots was in average further optimized for the vehicles than for the collision types, suggesting an overall higher level of dispersion along the road network for the latter. Furthermore, Table 4 indicates that accidents involving public buses presented the higher level of concentration at differential risk hotspots, followed by private buses, side collisions and rear-ends. Table 4 also points out that the procedure presented in this paper is specifically focused on the finding of microzones along the road network that are particularly dangerous for a collision or vehicle type, and not on the microzones where the collision or vehicle type is more concentrated (with higher accident counts), which usually associates with arterial and busy roads regardless of the collision or vehicle type.

Finally, an exploratory analysis of the coincidence of differential risk hotspots between collision and vehicle types was performed. Fig. 7 shows the percentage of road structure spanned by differential risk hotspots that is shared by each collision and vehicle type. This

Table 3
Summary statistics for all the sets of differential risk hotspots (denoted simply as hotspots in this table) that were determined as statistically significant by the Monte Carlo procedure (at the 0.05 level) for the collision and vehicle types considered, after the application of the detection methodology with $k = 1$ and $n = 40$. The table includes the number of differential risk hotspots that were found in each case, its total length (in kilometers), the number of traffic accidents of that type ($N_{type}^{hotspots}$) and in total ($N_{total}^{hotspots}$) that were recorded within the hotspots, the number of traffic accidents of that type ($N_{type}^{extended}$) and in total ($N_{total}^{extended}$) that were recorded within the hotspots or within an extension of 75 m around them, the proportion $N_{type}^{hotspots}/N_{total}^{hotspots}$ ($P_{type}^{hotspots}$), the proportion $N_{type}^{extended}/N_{total}^{extended}$ ($P_{type}^{extended}$) and the global proportion of the type of accident along the complete road network $p_{type}^{network}$ (also available in Table 1).

Type	Hotspots	Length	$N_{type}^{hotspots}$	$N_{total}^{hotspots}$	$N_{type}^{extended}$	$N_{total}^{extended}$	$P_{type}^{hotspots}$	$P_{type}^{extended}$	$P_{type}^{network}$
Crossing	48	2.97	251	416	660	1400	0.60	0.47	0.28
Fixed-object	13	0.63	27	54	106	373	0.50	0.28	0.14
Rear-end	43	5.15	409	866	823	2193	0.47	0.38	0.24
Run-off-road	19	1.13	15	108	39	537	0.14	0.07	0.02
Run-over	10	0.39	11	27	51	339	0.41	0.15	0.06
Side	30	4.42	237	867	410	2112	0.27	0.19	0.12
Bicycle	22	1.06	45	184	87	683	0.24	0.13	0.06
Lorry	11	0.98	34	185	63	543	0.18	0.12	0.04
Motorcycle	36	3.30	263	537	568	1485	0.49	0.38	0.26
Private bus	27	3.00	45	383	81	1429	0.12	0.06	0.02
Public bus	26	5.39	231	1047	339	1978	0.22	0.17	0.07
Van	15	0.77	31	100	73	423	0.31	0.17	0.08



(a)

Fig. 7. Percentage of coincident road length presented between the set of differential risk hotspots obtained for the collision and vehicle types considered after the application of the procedure with $k = 1$, $n = 40$ and $\sigma = 100$. This percentage is computed with respect to the collision or vehicle type whose set of hotspots spans over a larger number of meters (through the quotient between coincident road length and total length of the longer set of hotspots).

percentage is obtained in relation to the collision or vehicle type whose set of differential risk hotspots is longer. The highest percentages were found between side collisions and buses (both private and public), with a maximum percentage of coincidence close to 20% for side collision and public bus. Other remarkable but minor associations included crossing with motorcycles and rear-ends with motorcycles and private buses.

5. Conclusions

This paper has fully described a methodology that provides microzones of a road network that present a significant high risk for a specific type of accident. We called these microzones differential risk hotspots. The successful implementation of such methodology obviously requires that the road network itself is taken as the space where the traffic accidents are located (areal spaces are not accurate enough for this matter). Furthermore, it is recommended to fraction the road network into shorter pieces (lixels) to increase precision.

The use of a KDE-based technique which is typically used in case-control studies has rendered possible to estimate a probability of risk along the whole network for every type of collision and vehicle properly informed in the available dataset. From these estimates, a procedure that accounts for the disparity between each estimate and the mean proportion of the type of accident in the complete road network (that also excludes the parts of the network lacking a representative sample) has been designed. The election of the two parameters that define the procedure is vital to obtain a sensible number of differential risk hotspots.

The detection of microzones of a road network structure that present a differential risk for some type of collision or vehicle type should be a previous step to the definition of preventive measures that diminish its dangerousness. Depending on the resources available, the number of microzones of the road network that one would like to consider for further analysis and treatment could be very different. For this reason, the parameters of the algorithm could be chosen in a less restrictive way in order to find a larger number of microzones at which preventive measures might be implemented. Oppositely, if only a very reduced number of measures can be afforded, the values of the parameters should be increased and this naturally would lead to only a few microzones to be further studied.

In this regard, the differential risk hotspot detection procedure is complemented with the subsequent inclusion of a Monte Carlo technique that scores the importance of each hotspot. Hence, the hotspots presenting a statistical significance under a fixed threshold could be discarded or left aside from a first package of preventive measures.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

We thank José Serrano, Chief of Valencia Local Police Department, the *Unitat d'Atestats i Seguretat en el Transport* and the Statistics Office of Valencia for providing the data that was used in the study.

References

Anderson, J., Hernandez, S., 2017. Roadway classifications and the accident injury severities of heavy-vehicle drivers. Anal. Methods Accid. Res. 15, 17–28.
Anselin, L., 1995. Local indicators of spatial association-LISA. Geogr. Anal. 27 (2), 93–115.
Baddeley, A., 2010. Multivariate and marked point processes. In: Gelfand, A.E., Diggle, P., Guttorp, P., Fuentes, M. (Eds.), Handbook of Spatial Statistics. CRC Press, pp. 371–402 (Chapter 21).
Baddeley, A., Rubak, E., Turner, R., 2015. Spatial Point Patterns: Methodology and Applications With R. CRC Press.
Bíl, M., Andrášik, R., Janoška, Z., 2013. Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. Accid. Anal. Prevent. 55, 265–273.
Bivand, R., Piras, G., 2015. Comparing implementations of estimation methods for spatial econometrics. J. Stat. Soft. 63 (18), 1–36.
Bradley, J.V., 1978. Robustness? Br. J. Math. Stat. Psychol. 31 (2), 144–152.
Briz-Redón, Á., 2019a. DRHotNet: Differential Risk Hotspots in a Linear Network. R Package Version 1.0.
Briz-Redón, Á., 2019b. SpNetPrep: an R package using Shiny to facilitate spatial statistics on road networks. Res. Ideas Outcomes 5, e33521.
Chainey, S., Tompson, L., Uhlig, S., 2008. The utility of hotspot mapping for predicting spatial patterns of crime. Secur. J. 21 (1–2), 4–28.
Chang, L.-Y., Mannering, F., 1999. Analysis of injury severity and vehicle occupancy in truck-and non-truck-involved accidents. Accid. Anal. Prevent. 31 (5), 579–592.
Chang, L.-Y., Wang, H.-W., 2006. Analysis of traffic injury severity: an application of non-parametric classification tree techniques. Accid. Anal. Prevent. 38 (5), 1019–1027.
Coll, B., Moutari, S., Marshall, A.H., 2013. Hotspots identification and ranking for road safety improvement: an alternative approach. Accid. Anal. Prevent. 59, 604–617.
Copas, J., 1983. Plotting p against x. Appl. Stat. 25–31.
Dell'Acqua, G., Russo, F., Biancardo, S.A., 2013. Risk-type density diagrams by crash type on two-lane rural roads. J. Risk Res. 16 (10), 1297–1314.
Diggle, P., Zheng, P., Durr, P., 2005. Nonparametric estimation of spatial segregation in a multivariate point process: bovine tuberculosis in Cornwall, UK. J. R. Stat. Soc.: Ser. C (Appl. Stat.) 54 (3), 645–658.
Geedipally, S.R., Lord, D., 2010. Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of Poisson-gamma models. Accid. Anal. Prevent. 42 (4), 1273–1282.
Geedipally, S., Patil, S., Lord, D., 2010. Examination of methods to estimate crash counts by collision type. Transport. Res. Rec.: J. Transport. Res. Board 2165, 12–20.
Golob, T.F., Recker, W.W., Leonard, J.D., 1987. An analysis of the severity and incident duration of truck-involved freeway accidents. Accid. Anal. Prevent. 19 (5), 375–395.
Graul, C., 2016. leafletR: Interactive Web-Maps Based on the Leaflet JavaScript Library. R Package Version 0.4-0.
Harirforoush, H., Bellalite, L., 2016. A new integrated GIS-based analysis to detect hotspots: a case study of the city of Sherbrooke. Accid. Anal. Prevent.
Hosseinpour, M., Yahaya, A.S., Sadullah, A.F., 2014. Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes: case studies from Malaysian Federal Roads. Accid. Anal. Prevent. 62, 209–222.
Kahle, D., Wickham, H., 2013. ggmap: spatial visualization with ggplot2. R J. 5 (1), 144–161.
Kelsall, J.E., Diggle, P.J., 1995. Non-parametric estimation of spatial variation in relative

risk. Stat. Med. 14 (21–22), 2335–2342.
Kelsall, J.E., Diggle, P.J., 1998. Spatial variation in risk of disease: a nonparametric binary regression approach. J. R. Stat. Soc.: Ser. C (Appl. Stat.) 47 (4), 559–573.
Kelsall, J.E., Diggle, P.J., et al., 1995. Kernel estimation of relative risk. Bernoulli 1 (1–2), 3–16.
Kim, J.-K., Kim, S., Ulfarsson, G.F., Porrello, L.A., 2007. Bicyclist injury severities in bicycle-motor vehicle accidents. Accid. Anal. Prevent. 39 (2), 238–251.
McSwiggan, G., Baddeley, A., Nair, G., 2017. Kernel density estimation on a linear network. Scand. J. Stat. 44 (2), 324–345.
Moran, P.A., 1950a. Notes on continuous stochastic phenomena. Biometrika 37 (1/2), 17–23.
Moran, P.A., 1950b. A test for the serial independence of residuals. Biometrika 37 (1/2), 178–181.
Nie, K., Wang, Z., Du, Q., Ren, F., Tian, Q., 2015. A network-constrained integrated method for detecting spatial cluster and risk location of traffic crash: a case study from Wuhan, China. Sustainability 7 (3), 2662–2677.
Okabe, A., Sugihara, K., 2012. Spatial Analysis Along Networks: Statistical and Computational Methods. John Wiley & Sons.
Okabe, A., Satoh, T., Sugihara, K., 2009. A kernel density estimation method for networks, its computational method and a GIS-based tool. Int. J. Geogr. Inf. Sci. 23 (1), 7–32.
OpenStreetMap Contributors Planet Dump, 2017. Retrieved from: https://www.openstreetmap.org. 2017.
R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Robey, R.R., Barcikowski, R.S., 1992. Type I error and the number of iterations in Monte Carlo studies of robustness. Br. J. Math. Stat. Psychol. 45 (2), 283–288.
Savolainen, P., Mannering, F., 2007. Probabilistic models of motorcyclists' injury severities in single-and multi-vehicle crashes. Accid. Anal. Prevent. 39 (5), 955–963.
Serra, L., Juan, P., Varga, D., Mateu, J., Saez, M., 2013. Spatial pattern modelling of wildfires in Catalonia. Spain 2004–2008. Environ. Model. Softw. 40, 235–244.
Shankar, V., Mannering, F., 1996. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. J. Safe. Res. 27 (3), 183–194.
Silverman, B.W., 2018. Density Estimation for Statistics and Data Analysis. Routledge.
Thakali, L., Kwon, T.J., Fu, L., 2015. Identification of crash hotspots using kernel density estimation and kriging methods: a comparison. J. Mod. Transport. 23 (2), 93–106.
Van Patten, I.T., McKeldin-Coner, J., Cox, D., 2009. A microspatial analysis of robbery: prospective hot spotting in a small city. Crime Map.: J. Res. Pract. 1 (1), 7–32.
Walker, I., 2007. Drivers overtaking bicyclists: objective data on the effects of riding position, helmet use, vehicle type and apparent gender. Accid. Anal. Prevent. 39 (2), 417–425.
Wang, K., Ivan, J.N., Ravishanker, N., Jackson, E., 2017. Multivariate Poisson lognormal modeling of crashes by type and severity on rural two lane highways. Accid. Anal. Prevent. 99, 6–19.
Xie, Z., Yan, J., 2008. Kernel density estimation of traffic accidents in a network space. Comput. Environ. Uban Syst. 32 (5), 396–406.
Xie, Z., Yan, J., 2013. Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. J. Transp. Geogr. 31, 64–71.
Ye, X., Pendyala, R.M., Washington, S.P., Konduri, K., Oh, J., 2009. A simultaneous equations model of crash frequency by collision type for rural intersections. Safe. Sci. 47 (3), 443–452.