An Acad Bras Cienc (2022) 94(Suppl.4): e20210727 DOI 10.1590/0001-3765202220210727
Anais da Academia Brasileira de Ciências | *Annals of the Brazilian Academy of Sciences*
Printed ISSN 0001-3765 I Online ISSN 1678-2690
www.scielo.br/aabc | www.fb.com/aabcjournal

GEOSCIENCES

Using Self-Organizing Maps to find spatial relationships between wildlife-vehicle crashes and land use classes

LARISSA S. TSUDA, CLEYTON C. CARNEIRO & JOSÉ ALBERTO QUINTANILHA

Abstract: The construction and expansion of roads cause significant impacts on the environment. The main potential impacts to biotic environment are vegetation suppression, reduction of the amount and composition of animal distribution due to forest fragmentation and increasing risks of animal (domestic and wildlife) vehicle collisions. The objective of this work was to establish a relationship between the different spatial patterns in wildlife-vehicle crash, by using spatial analysis and machine learning tools. Self-Organizing Maps (SOM), an artificial neural network (ANN), was selected to reorganize the multi-dimensional data according to the similarity between them. The results of the spatial pattern analysis were important to perceive that the point data pattern varies from an animal type to another. The events occur spatially clustered and are not uniformly distributed along the highway. SOM was able to analyze the relationship between multiple variables, linear and non-linear, such as ecological data, and established distinct spatial patterns per each animal type. In the studied area, most of the wildlife was run over very close to forest area and water bodies, and not so close to sugarcane fields, forestry and built environment. A considerable part of the wildlife-vehicle collisions occurred in areas with diverse landscape.

Key words: Accident prevention, artificial neural networks, geographic information systems, machine learning, road safety, wildlife-vehicle crash.

INTRODUCTION

Globally, most terrestrial wildlife deaths are caused by wildlife-vehicle collisions (Forman 1998, Freitas et al. 2010, Teixeira et al. 2013). Aside from their direct impact on biodiversity, wildlife-vehicle collisions affect the safety and assets of road users (Huijser et al. 2013). Therefore, it is necessary to investigate the locations and circumstances under which wildlife-vehicle collisions occur and to identify the spatial patterns connecting them. This information would inform not only the efforts of infrastructure planners, road safety specialists, and managers of wildlife populations, but also the development of stable mitigation measures aimed at improving driver and animal safety and minimizing economic losses.

Litvaitis & Tash (2008) affirm that research to understand the factors that contribute to wildlife-vehicle collisions (WVC) can be partitioned into several major themes, including (i) characteristics associated with roadkill hot spots, (ii) identification of road-density thresholds that limit wildlife populations, and (iii) species specific models of vehicle collision rates that incorporate information on roads (e.g., proximity, width, and traffic volume) and animal movements. Results revealed important species-specific differences, with traffic volume and rate of movement by candidate species having the greatest influence on collision rates.

Intending to summarize empirical WVC findings to facilitate the application of this knowledge to the planning, and design of mitigation strategies on roads, Gunson et al. (2011) conducted a review restricted to manuscripts that used generalized linear models to statistically determine the influence that numerous explanatory predictors have on the location of WVCs. They conclude that WVCs exhibit clustering on roads, which is attributed to specific landscape and road-related factors.

In a similar spatial analysis approach to this our paper, Morelle et al. (2013) report data on WVC in Wallonia, southern Belgium for wild boar, roe deer, red deer, and red fox, clustering the accidents for all these species, mapping via Kernel density analysis. The Authors also performed a temporal analysis which was not developed by us given the lack of data for this type of analysis.

Sáenz-de-Santa-María & Tellería (2015) review official, unpublished information on WVC provided by the Spanish authorities to assess the main features and geographic distribution of this wildlife human interaction. They use the reported vehicle collisions to explore: (a) the identity, conservation status, and biological traits of species involved in most WVC; (b) the distribution of those areas within the country where these collisions are more likely to occur; and (c) the economic and human costs (injured persons) of this wildlife-human interaction.

Bíl et al. (2016), present an objective method for hotspot identification that can be used for animal-vehicle collisions (AVC) data, using the Kernel density estimation – KDE and determining the significance level of hotspots. According to the Authors, the prioritization of hotspots allows a transportation manager to effectively allocate resources to a feasible number of identified hotspots.

In Pagany (2020) a review of WVC is presented and identify factors such as the proximity to forest, a gentle topography with sparsely curves, street width, and seasonal differences are common denominators for WVCs - independent of the species -, while traffic volume, the distance to urban areas, or road accompanying infrastructure are not assignable influencing or non-influencing factors.

According to Shilling et al. (2020): "Preventing WVC begins with recording locations of conflict, such as vehicle crashes, animal carcasses (roadkill), or animal behavior around roads, such as avoidance of roads or crossing-behavior". They developed a web-systems for reporting wildlife-vehicle-conflict reviews and provided recommendations for future WVC reporting systems and guidelines for the management of road networks.

Schwartz et al. (2020) affirm "The number of WVCs has an obvious value in estimating the direct effects of roads on wildlife, i.e. mortality due to vehicle collisions". They provide a review through a series of case studies contributing to the advancement of knowledge in species distributions, population dynamics, and animal behavior, as well as informing us about the health of the species and of the environment. Propose that monitoring roadkill facilitates five critical areas of ecological study: (1) monitoring of roadkill numbers, (2) monitoring of population trends, (3) mapping of native and invasive species distributions, (4) animal behavior, and (5) monitoring of contaminants and disease. Our paper agrees with "monitoring of roadkill numbers" with the proposal to provide guidelines for transportation planning.

Valerio et al. (2021) identify spatio-temporal trends of roadkill occurrence using citizen science data from one of the most urbanized and biodiversity-rich regions of Italy. Temporal trends were analyzed using generalized additive models, while landscape patterns were assessed by identifying significant thresholds over land cover gradients, related to increases in relative roadkill abundance, by employing threshold indicator taxa analysis. They develop a map of potential roadkill risk that could assist in planning the placement of mitigation measures. Citizen science contributions from highly populated areas allowed data collection over a large area and a dense road network, and also directly led to the evaluation of management decisional options.

Studies discussing the results of primary surveys of wildlife-vehicle collisions (Rowden et al. 2008), focus on the specific species of animals involved in such collisions (Taylor et al. 2002, Huijser & Bergers 2000), evaluating the efficiency of wildlife crossings (Glista et al. 2009), estimating the costs generated by wildlife-vehicle collisions and implementing mitigating measures (Huijser et al. 2013, 2009), and using spatial analysis methods to understand the spatial patterns of wildlife-vehicle collisions (Barthelmess 2014, Clevenger et al. 2003, De Freitas et al. 2015, Gunson et al. 2011), among others, have been conducted. However, these studies are relatively recent. Further research and a greater amount of data are needed to improve the identification of the species involved in collisions with vehicles, evaluate the consistency of wildlife crossings, and identify spatial patterns in wildlife-vehicle collisions.

Self-organizing maps (SOMs) are tools based on artificial neural networks (ANNs) and are used to analyze and visualize multidimensional data. They reorganize an N-dimensional dataset, where N is the number of variables involved in a two-dimensional (2D) map. Based on sequences of competition and collaboration, best matching units (BMUs) are chosen to represent sample groups according to the similarity of their relationships. A SOM is an ideal method for analyzing complex data, as it can be used to extract linear and non-linear relationships from a data set, and can be effectively applied to classification and association (Park et al. 2003), for example, Kussul et al. (2017) used a SOM to restore missing values in low-resolution satellite images using neural coefficient weights.

The main objective of this study was to establish a relationship between the different spatial patterns in wildlife-vehicle collisions by using spatial analysis and machine learning tools to understand the connection between wildlife-related collisions, the animal species involved, and the variables that represent land cover and road characterization features. To this end, a SOM algorithm was used to extract the relationships between wildlife species involved in collisions and the spatial characteristics that contextualize crash events. These characteristics included proximity to and the relative area of forest formations, water bodies, silviculture, and built-up areas, as well as the maximum speed allowed on the road, traffic volume, landscape diversity, and time of the day when the crash occurred.

MATERIALS AND METHODS

The study area is a 207 km stretch of toll highway intercepting 14 municipalities that are located in the western part of the state of São Paulo, Brazil. The land cover in the study area consists of pastures and anthropogenic areas, sugarcane fields, silviculture, water bodies, and built-up areas (FBDS 2017). The study area is located in the western central plateau of Brazil and is characterized by wide and low hills, which range from being minimally eroded to being flat, with valleys that are not deep and a low drainage density (Ross & Moroz 1997).

This study used secondary data collected by public agencies—the Secretaria do Meio

Ambiente do Estado de São Paulo (SMA), Fundação Brasileira para o Desenvolvimento Sustentável (FBDS), Instituto Nacional de Pesquisas Espaciais (INPE), and the highway management company. Primary data were not used in this study. The data collected were related to collisions involving animals, land cover, and land cover in the areas around the highway, as shown in Figure 1, and the characteristics of the highway itself (Table I). Animal-related collision data were obtained from CETESB, public process no. 13716/2001 and no. 154/2011. Figure 2 shows a flowchart summarizing the methodology.

A total of 1469 wildlife-vehicle collisions involving 16 types of animals that occurred between 2011 and 2015 (Table II) were mapped. Mammals, the animal group most reported, and potentially endangered animals were studied further. This is because wildlife-vehicle collisions are usually monitored from a motor vehicle and because medium or large mammals (those over 1 kg) are easier to see from such vehicles than smaller animals, such as most amphibians, some mammals, and reptiles (Glista et al. 2008). Furthermore, medium and large animals are more likely to be involved in accidents that cause material damage and present a fatal risk to highway users (Huijser et al. 2013, Bueno et al. 2013).

In addition to the wildlife-vehicle collision locations, this study examined 32 variables (Table II) related to the characteristics of the road and the land use and land coverage of the surrounding area. The variables were chosen based on the data available at the time the research was conducted and the results of previous studies (Barthelmess 2014, Gunson et al. 2011, Bueno et al. 2013, Carneiro et al. 2012). The decision to use the SOM algorithm was justified by the high number of variables.

The distances from each collision point to the nearest forest formation, water bodies, silviculture, sugarcane fields, and built-up area were calculated, and circular buffers were also considered around each crash point. The relative areas of the land cover classes within each of these buffers were measured as a percentage of the total area. The variables

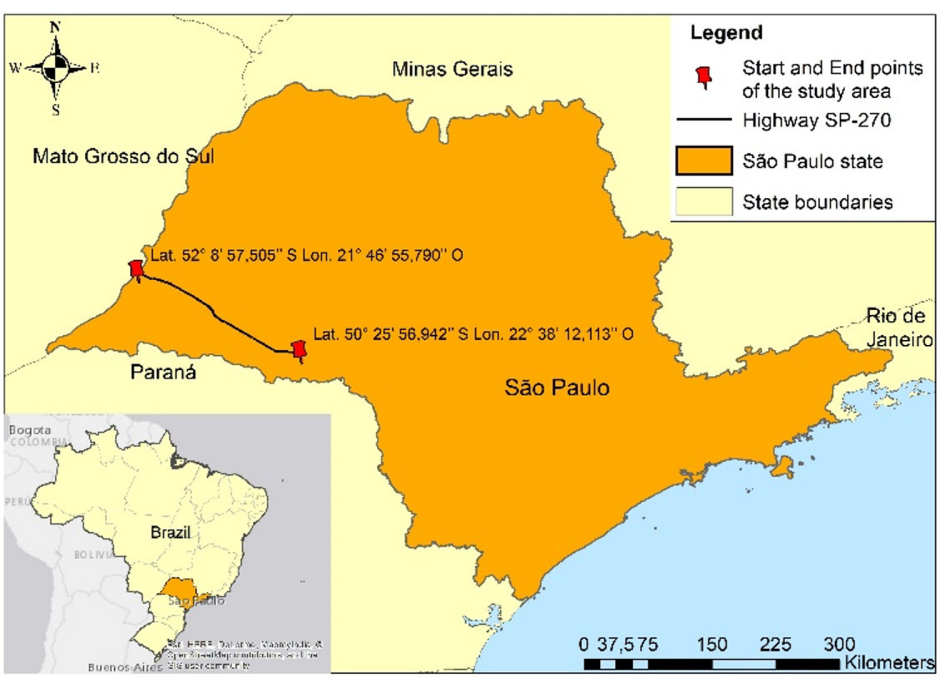

Figure 1. Map of the study area.

related to the relative area, landscape diversity, and total length of rivers in the surroundings were measured in three radius sizes, i.e., 500 m, 1 km, and 5 km, resulting in analysis areas of 0.78 km², 3.14 km², and 78.53 km², respectively. This was done to identify the radius within which the greatest number of collisions with animals occurred.

The Shannon–Wiener Diversity index (1949) was used to measure the number of species within a community (Spellerberg & Fedor 2003). In this study, the index was used to identify only the number of different land cover types in one parcel. This index does not differentiate types of land cover by biological importance; rather, it considers all types of coverage as equally important. For instance, forest and sugarcane fields received the same weight even though their significance to different animal species, which was not considered in our analysis, is different. It makes sense since we are identifying the hot spot occurrences places and then, trying to associate them with land cover, not the opposite.

The species involved in the collisions were originally codified as categorical variables; however, for SOM to be applied, the variables must be converted into numerical values. Binary code, which transforms each categorical attribute into a set of binary attributes (Hsu 2006), was used to convert each categorical value into a binary attribute. As a result, 16 columns were created, one for each species involved in the collisions (Table II), and these were populated with a 1 if an animal of a given species was run over (involved in a crash) and a 0 if an animal of a given species was not run over. In addition to the animal columns, 32 variables related to the characteristics of the road and land cover (Table III) were added as columns. The SOM pre-training database totaled 1469 rows, each of which represented one sample (one collision event), and 48 columns, each representing a variable. To ensure that the variables contributed the same

Table I. List of Secondary Spatial Daa used for the research work.

Data type		Geographic data type	Year	Scale	Responsible agency	Acess at
forest formation	land use	vector/polygon	2013	1:20,000	FBDS	http://www.fbds.org.br
water body			2013			
urban area			2013			
silviculture			2013			
river		vector/line	2013			
sugarcane		vector/polygon	2013	1:50,000	INPE	http://www.dsr.inpe.br/laf/canasat
elevation		raster/pixel	2013	1:50,000	SMA	http://datageo.ambiente.sp.gov.br
wildlife-vehicle collision		vector/point	2011-2015	local	concessionary[1]	Cetesb process nº 13716/2001 e nº 154/2011
highway		vector/line	2014	-	OpenStreetMap	https://www.openstreetmap.org

FBDS - Fundação Brasileira para o Desenvolvimento Sustentável; INPE - Instituto Nacional de Pesquisa Espaciais; SMA - Secretaria do Meio Ambiente do Estado de São Paulo; Cetesb - Companhia Ambiental do Estado de São Paulo.
[1]The fauna running over data were obtained through public processes Cetesb nº 13.716/2001 and nº 154/2011. Organization: Larissa S. Tsuda.

weight to the analysis, they were normalized based on their variance before the SOM analysis was performed.

An exploratory SOM analysis was carried out to filter the variables that, according to the presented correlation, were related to wildlife-vehicle collisions (Figure 3). The elimination of variables that had a low correlation with the latter decreased map dimensionality and allowed the main relationships between variables to gain greater prominence in the SOM. The exploratory analysis was completed in two stages.

In the first stage, exploratory analysis 1 (EA1) only considered 21 variables that were repeated at more than one scale (radius of 500 m, 1 km, and 5 km) and 16 variables related to wildlife-vehicle collisions. Only the scale with the highest correlation to the number of wildlife vehicle collisions was selected per variable. Exploratory Analysis 2 (EA2) considered 16 animal species and 18 variables related to road characteristics and land cover, all of which were eliminated in EA1. It identified and eliminated variables with low correlations ($r < 0.099$) with wildlife vehicle collisions. In the second stage, a definitive analysis (DA) was developed. This included 31 variables, of which 16 were animal species and the other 15 were related to land cover and road characteristics.

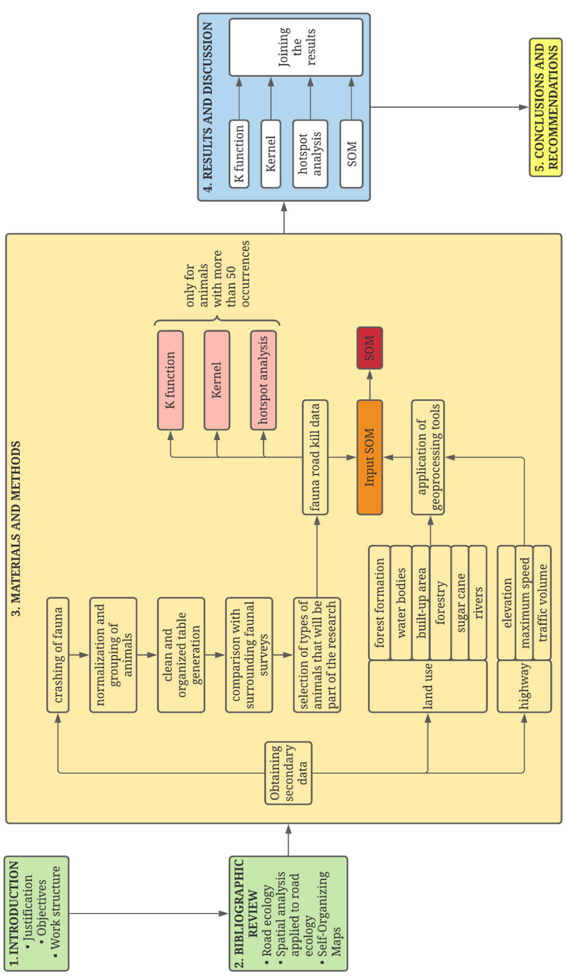

Figure 2. Flowchart summarizing the adopted methodology.

In other similar studies, two common rules were followed when choosing the map size, that is, the number of map units that make up the matrix. Vesanto et al. (2000) recommended that the number of map units in the matrix be approximately $5\sqrt{N}$, where N is the total number of samples in each variable. Kohonen (2013) stated that it was not possible to estimate the exact map size beforehand and that this number should be determined based on trial and error by comparing the results obtained from each trial and then selecting the size. Our

Table II. Animals (Variables) considered in the study.

N	Variable	Unit
1	Presence of run over tapir	Number of animals
2	Presence of run over canid	
3	Presence of run over capybara	
4	Presence of run over deer	
5	Presence of a trampled opossum	
6	Presence of run over raccoon	
7	Presence of run over ocelot	
8	Presence of run over hare	
9	Presence of a trampled maned wolf	
10	Presence of run over monkey	
11	Presence of a trampled puma	
12	Presence of run over hedgehog	
13	Presence of a trampled coati	
14	Presence of a trampled giant anteater	
15	Presence of a trampled anteater	
16	Presence of a trampled armadillo	

dataset consisted of 1469 samples. The map size calculated using the method indicated by Vesanto et al. (2000) was approximately 191.65 units, which resulted in a 14 × 14 map unit matrix. Several maps were created for other sizes, and it was noted that the larger maps tended to have smaller quantization errors, which is the average distance between each sample and its BMU. The topographical error, which is used to measure topology preservation, represents the proportion of all samples for which the first and second BMUs are not adjacent (Céréghino & Park 2009). According to Compin & Céréghino (2007), a map that is too small might not efficiently explain some of the important differences that should be detected; however, if the map is too large, the differences observed would be very small. In other words, the larger the map, the lower the number of samples per neuron, and the fewer the samples that can be grouped based on similarity.

The size of the map was set as 14 × 14 units with toroidal projection and hexagonal cells. The algorithm training commenced with a random

Table III. Land use types and road characteristics (Variables) considered in the study.

N	Variable	Unit
17	Distance from the nearest forest formation	
18	Distance from the nearest forestry area	
19	Distance from the nearest water body	m
20	Distance from the nearest build-up area	
21	Distance from the nearest sugarcane growing area	
22	Percentage of forest formation in the 500m buffer	
23	Percentage of forestry area in the 500m buffer	
24	Percentage of water bodies in the 500m buffer	
25	Percentage of built-up area in the 500m buffer	
26	Percentage of sugarcane in the 500m buffer	
27	Percentage of forest formation in the 1km buffer	
28	Percentage of forestry area in the 1km buffer	
29	Percentage of water bodies in the 1km buffer	%
30	Percentage of built-up area in the 1km buffer	
31	Percentage of sugarcane in the 1km buffer	
32	Percentage of forest formation in the 5km buffer	
33	Percentage of forestry area in the 5km buffer	
34	Percentage of water bodies in the 5km buffer	
35	Percentage of built-up area in the 5km buffer	
36	Percentage of sugarcane in the 5km buffer	
37	Total length of linear rivers withn the 500m buffer	
38	Total length of linear rivers withn the 1km buffer	
39	Total length of linear rivers withn the 5km buffer	
40	Shannon-Wiener Diversity Index for land use classes within the 500m buffer	
41	Shannon-Wiener Diversity Index for land use classes within the 1km buffer	
42	Shannon-Wiener Diversity Index for land use classes within the 5km buffer	
43	Traffic volume	vehicles/year
44	Maximum speed allowed	km/h
45	Elevation (altitude)	m
46	Time of the day: Morning	
47	Time of the day: Afternoon	-
48	Time of the day: Night	

sample for which the closest neuron was selected. This process was repeated for all the samples, and the training was completed in two stages. The same training parameters were used for EA1, EA2, and DA. The data generated during DA were displayed using 2D maps for each of the variables. In these maps, the color scale shows the neurons that have a greater variable contribution to the analysis. Furthermore, neuron clustering was completed using a U-matrix based on k-means and the Davies–Bouldin index (DBI) (Davies & Bouldin 1979). It displays the results more comprehensively and quantitatively.

RESULTS

No relationship was observed between the elevation data and maximum allowable vehicle speed, and wildlife-vehicle collisions. The maximum allowable vehicle speed was the same as that provided on road signs, and it did not vary significantly over the study area. The elevation data did not change significantly, given that the terrain of the study area was predominantly flat

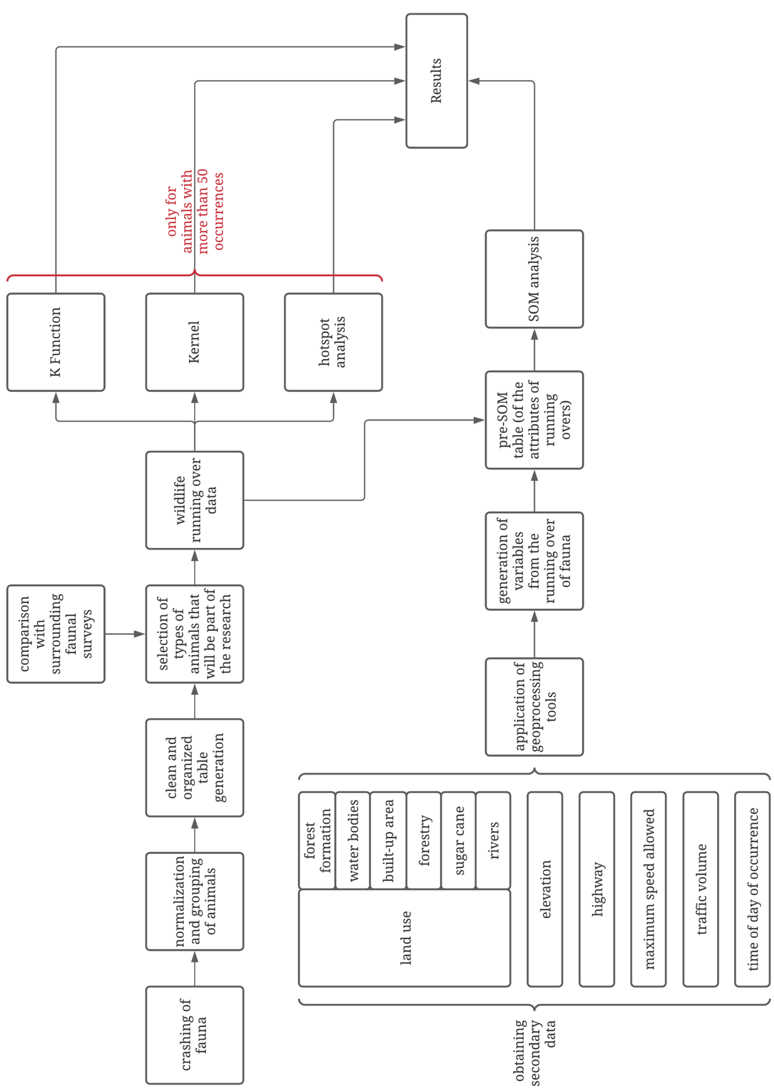

Figure 3. Flowchart of the exploratory SOM analysis.

and composed of wide and low hills. Other data that could be related more to wildlife-vehicle collisions include the longitudinal profile of the roads and the actual speed of vehicles involved in the collisions. Only data available at the time the research was being conducted were used. Data on the longitudinal profile of the road and actual vehicle speeds were more difficult to obtain.

The following variables were selected from EA1: percentage of forest formation within a buffer of 500 m, percentage of water bodies within a buffer of 500 m, percentage of silviculture within a buffer of 5 km, percentage of the built-up area within a buffer of 5 km, percentage of sugarcane fields within a buffer of 500 m, total length of linear rivers within a buffer of 500 m, and the Shannon–Wiener Diversity index for types of land cover within a 1 km buffer.

EA2 generated a 14 × 14 (196) unit map, and the samples were grouped into 127 BMUs. The final quantization error was 3.35. The eliminated variables were elevation ($r < 0.07$), maximum speed of the vehicle ($r < 0.087$), and time of day: afternoon ($r < 0.034$). The DA generated a 14 × 14 (196) unit map, and the samples were grouped into 126 BMUs. The final quantization error was 3.02.

Two-dimensional maps for each variable are shown in Figure 4. It should be noted that, although most of the studied animals are found in specific regions, a capybara can be found in all regions; the maps of the different distances are quite complementary among them (justifying their choice), and the spatial behavior of the selected variables. The averages for the variables for land cover and highway characterization by animal species are shown in Table IV. The data on the distances from various land cover types (Table IV) confirmed that, in general, the animals were run over 315 m and 311 m, on average, from forest fragments and water bodies, respectively. In comparison, wildlife-vehicle collisions occurred 3.4 km, 2.2 km, and 5.8 km from the closest sugarcane fields, silviculture, and built-up areas, respectively.

A second clustering process was applied to refine the initial process that generated the 2D variable map. This second clustering was optional, and its purpose was to present the results more clearly. The DBI was used to determine the ideal number of clusters for the dataset. To reach a statistically consistent ideal number of clusters, the DBI was applied 70 times, and the ideal mode was selected. This resulted in the creation of 15 clusters (Figure 5).

The animals involved in the highest number of collisions, capybaras (41%) and armadillos (18%), were found in large numbers in multiple clusters (Table V). No unique spatial patterns were observed for these species. Hares, canids, and collared anteaters, which were also involved in several recorded collisions, were grouped into more than one cluster. However, they were predominantly in a single cluster, which indicated a more defined pattern. The other species, which were involved in 35 collisions or fewer, were grouped into single clusters. This indicated a more defined special pattern. In general, animals were run over when they were closer to forest formations, i.e., at an average distance of 315.3 m from the forest formations. Canids and maned wolves were involved in collisions further away from forests or in areas with low-density forests. Lowland tapirs, ocelots, pumas, and monkeys were involved in collisions that occurred closer to forests and in high-density forests.

These results are consistent with the literature since, according to Silveira et al (2010), some species search for food and water in dry seasons, which increases the contact with roads. In the studies of Abra (2019) and Abra et al. (2021), road kills are improved, during dry

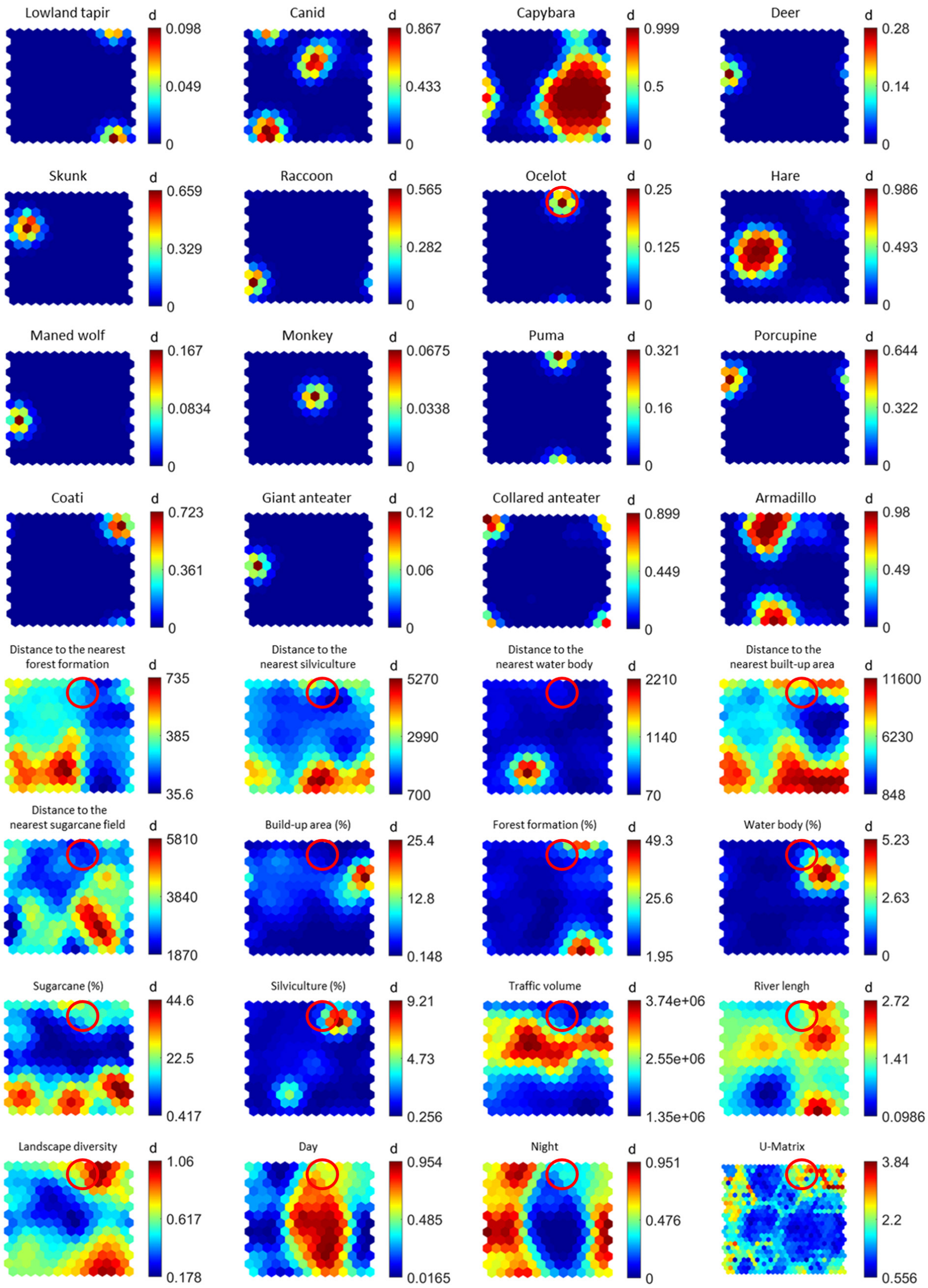

Figure 4. **Component plots showing two-dimensional maps for the variables used in the SOM analysis. Each map shows the activation of neurons based on the values of the analyzed variables. In this way, warm colors show the high values of the variables, and cool colors the low values. The last map corresponds to the U-Matrix, where it is possible to observe the similarity between neighboring samples. In this case, cool colors correspond to the high similarity between neighboring samples, while warm colors denote greater dissimilarity. The red circles exemplify the common characteristics of running over the ocelot: low distance to the nearest forest formation; regions close to sugarcane and silviculture; low traffic volume; high landscape diversity; at the end of the day or the beginning of the evening; among others.**

seasons, for some species, including manned-wolf and ocelot. The maned wolf, the largest canid in South America, is subject to many threats, including road kills and inhabiting habitats such as woodland with an open canopy, mixed forest and grassland, and wet fields (https://seaworld.org/animals/facts/mammals/maned-wolf/). Lowland tapir, also known as the Brazilian tapir, lives in the rain forests of South America (https://tapirs.org/wp-content/uploads/2018/06/TAPIR-TRACKS-A-Curriculum-Guide-for-Educators.pdf). Ocelots have been recorded in a great variety of habitats, from heavily logged and fragmented forests, to early and late successional forests, the outskirts of major cities and towns, disturbed scrub/

Table IV. Animals involved incrashes and the average for land cover and highway characterization variables.

Animals	Distance to the nearest forest formation	Forest formation percentage in a 500 m buffer	Distance to the nearest water body	Water body percentage in a 500 m buffer	Distance to the nearest silviculture	Silviculture percentage in a 5 km buffer	Distance to the nearest built-up area	Built-up area percentage in a 5 km buffer	Distance to the nearest sugarcane field	Surgarcane field percentage in a 500 m buffer	Traffic volume	Diversity of land use types in a 1 km buffer	Total length of the linear rivers in a 500 m buffer
Armadillo	399,6	5,7	554,7	0,3	2.287,8	1,6	6.040,9	3,09	2.724,4	16,0	2.157.373	0,5	1,0
Canid	406,4	5,5	342,2	0,3	2.215,7	1,3	6.276,7	2,69	2.978,8	20,2	2.094.463	0,6	1,1
Capybara	245,3	9,2	204,9	0,9	2.386,3	1,0	5.720,7	3,91	4.128,6	15,6	2.275.000	0,6	1,5
Coati	293,6	10,4	167,3	1,2	2.734,0	2,0	7.594,1	2,28	2.804,4	17,8	2.044.316	0,7	1,7
Collared anteater	383,0	11,2	283,9	0,4	2.240,2	1,4	7.507,7	1,83	3.439,8	17,1	1.988.907	0,6	1,3
Deer	206,6	11,3	512,4	0,3	1.969,3	0,6	5.361,6	3,22	2.140,1	9,6	2.710.043	0,6	1,2
Giant anteater	267,7	16,5	198,4	0,0	3.127,9	0,2	6.755,8	0,61	4.272,8	29,6	2.167.683	0,5	1,9
Hare	339,2	5,9	327,6	0,3	1.769,2	1,1	5.058,3	4,56	3.209,3	8,2	2.709.195	0,4	1,4
Lowland tapir	18,4	44,1	155,1	0,0	2.236,4	1,3	9.617,3	0,00	5.596,2	31,5	1.314.221	1,2	1,0
Maned wolf	479,4	1,4	123,7	0,0	2.759,8	0,5	8.662,8	0,54	1.460,9	10,4	2.780.044	0,4	1,8
Monkey	89,5	19,0	147,1	0,2	3.363,2	0,7	6.230,0	0,94	423,4	9,4	2.659.721	0,6	1,9
Ocelot	248,1	9,8	291,3	0,6	1.950,0	2,3	4.691,8	2,06	3.022,7	19,8	2.211.378	0,7	1,1
Porcupine	324,9	3,7	247,0	0,5	2.148,7	0,8	3.919,1	5,89	3.711,1	19,1	2.726.528	0,5	1,3
Puma	209,6	21,0	439,0	0,1	1.683,0	1,1	6.970,0	1,62	1.969,8	18,8	1.870.553	0,7	1,1
Skunk	277,6	7,8	304,0	0,1	1.789,9	0,5	4.629,4	5,38	2.262,4	15,3	2.511.563	0,5	1,5
South American raccoon	387,9	6,3	310,2	0,5	3.886,9	0,4	8.574,1	5,67	702,1	30,8	2.333.978	0,5	1,2
Total	315,3	7,9	311,6	0,6	2.259,9	1,2	5.874,2	3,62	3.407,7	15,6	2.293.886	0,6	1,3
	meters	**%**	**meters**	%	meters	%	meters	%	meters	%	vehicles/ year	-	km

woodland Savannah, and agricultural areas (de Oliveira et al. 2010). Results from Azevedo et al. (2021) revealed that the puma habitat has an association with forest vegetation, followed by pasture with shrubs. Road kills are important factors linked to the reduction of the cougar population in several areas (Azevedo et al. 2013, Benatti 2021). Titi monkeys occur in the Brazilian states of Rio de Janeiro, São Paulo, and Minas Gerais. Nowadays their population is restricted to forest patches within a highly fragmented

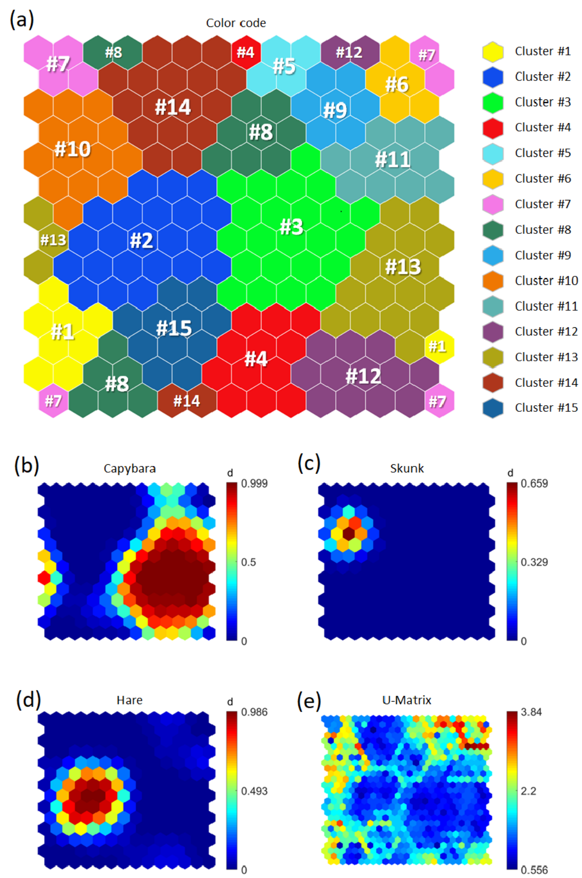

Figure 5. (a) Clustering of 15 regions with high similarity obtained based on the (e) U-Matrix and three examples for (b) capybara, (c) skunk, and (d) hare, respectively. In these three examples, the red color indicates a high concentration of accidents.

landscape, where local extinction is a constant threat (Trevelin et al. 2007).

In most occurrences, animals were run over when they were closer to water bodies. Lowland tapirs, South American coatis, maned wolves, and monkeys were involved in collisions that occurred very close to water bodies (Table IV). No species were run over far from the water.

The animals, in general, were run over far from silviculture, at an average distance of 2.2 km. No species were predominantly run over very close to silviculture. Raccoon collisions (cluster 1), 18.5% of the capybaras' collisions (clusters 4 and 12), and 14% of the armadillos (cluster 4) collisions occurred farther away from silviculture in areas with low silviculture density.

Sugarcane occupies a significant amount of land in the study area. However, collisions occurred far away from the sugarcane fields, at an average distance of 3.4 km; lowland tapirs were run over farther from sugarcane fields than the average distance; South American raccoons and monkeys were run over closer to sugarcane fields than the average distance. Fourteen percent of the armadillos' collisions (cluster 4), 47% of the capybaras' collisions (clusters 4 and 13), and 92% of the canid collisions occurred in areas with high densities of sugarcane fields.

In most cases, collisions with animals occurred far from built-up areas, at an average distance of 5.8 km. No species was run over very close to built-up areas in general, although in highly dense regions, South American raccoons, hares, 28.6% of the capybaras, deer, skunks, and porcupines were run over. Of all the animals studied, a small proportion (2.38%, cluster 9), consisting of capybaras, canids, hares, collared anteaters, and armadillos, were run over very close to silviculture, at an average distance of 242 m. These areas have a very diverse landscape, including a high proportion of water bodies (rivers and ponds) and sugarcane fields, and are near forest fragments.

Another small proportion (4.83%, cluster 11) of the animals studied, particularly capybaras, canids, hares, porcupines, and collared anteaters, was run over in areas with a high density of water bodies, a high density of built-up areas, and a high traffic volume. In these areas, the landscape is composed of forest fragments, rivers, dams, lagoons, and urban patches. Lowland tapirs and South American coatis were run over in areas with high landscape diversity; however, hares and maned wolves were run over in areas with low landscape diversity. Furthermore, the latter, along with monkeys,

Table V. **Averages for distances to different types of land cover and highway characterization, per cluster, in the definitive analysis of the self-ordering map.**

Cluster	Animals	Distance to the nearest forest formation (m)	Forest formation percentage in a 500 m buffer	Distance to the nearest water body (m)	Water body percentage in a 500 m buffer	Distance to the nearest silviculture (km)	Silviculture percentage in a 5 km buffer	Distance to the nearest built-up area (km)	Built-up area percentage in a 5 km buffer	Distance to the nearest sugarcane field (km)	Surgarcane field percentage in a 500 m buffer	Traffic volume (10^6 vehicles/year)	Diversity of land use types in a 1 km buffer	Total river lengh in a 500 m buffer (km)	Time of day: morning	Time of day: night
1	South American raccoon (100%),capybara (0,5%)	662,66	5,52	274,18	0,56	3,43	0,38	9,71	5,01	0,71	29,62	2,2	0,55	1,34	8	15
2	hare (87,6%), maned wolf (100%)	330,29	4,25	222,90	0,16	1,71	0,90	4,95	4,07	3,24	7,36	2,8	0,40	1,44	67	108
3	capybara (28,6%), macaco (100%)	299,92	4,87	151,95	0,47	1,71	1,11	4,57	4,25	4,13	0,74	2,8	0,32	1,45	161	0
4	capybara (9%),hare (1%), armadillo (14,7%)	514,97	4,85	366,38	0,25	5,10	0,18	11,20	0,07	2,92	38,06	1,5	0,70	0,89	78	10
5	ocelot (100%), puma(100%)	226,93	15,95	372,54	0,30	1,80	1,68	5,94	1,82	2,44	19,25	2,0	0,73	1,14	9	10
6	coati (100%)	293,56	10,37	167,31	1,20	2,73	2,04	7,59	2,28	2,80	17,84	2,0	0,74	1,70	13	14
7	collared anteater (94,6%)	386,58	11,53	256,92	0,31	2,31	1,11	7,80	1,35	3,26	17,37	2,0	0,60	1,28	28	39
8	canid (88,2%)	442,62	3,99	300,47	0,13	2,34	0,67	6,57	2,19	3,15	20,64	2,1	0,52	1,06	53	59
9	canid (4,6%), capybara (2,5%), hare (1%), collared anteater (2,7%) armadillo (3,4%)	180,74	4,93	173,61	2,73	0,24	12,12	2,30	0,12	2,18	23,21	1,3	1,14	1,44	19	12
10	deer(100%), skunk (100%), porcupine (96,9%), giant anteater (100%)	287,88	7,11	300,78	0,29	1,98	0,61	4,60	4,46	2,96	17,02	2,6	0,50	1,40	22	47
11	canid (2,6%), capybara (9,6%), hare (2,4%), porcupine (3,1%), collared anteater (1,4%), armadillo (0,8%)	68,37	12,47	63,36	5,15	1,90	0,18	0,74	19,00	4,22	2,67	3,6	0,65	2,41	30	32
12	lowland tapir (100%), canid (2%), capybara (9,5%), hare (2,9%), armadillo (1,9%)	27,68	43,27	85,23	0,28	3,82	0,67	12,06	0,13	5,13	17,73	1,4	1,02	2,46	44	21
13	capybara (38,4%)	220,33	5,80	201,23	0,24	2,31	0,58	5,10	2,95	3,60	24,86	2,0	0,59	1,24	35	167
14	armadillo (67,2%)	316,17	5,11	248,05	0,16	1,70	0,72	5,27	4,08	2,88	10,11	2,5	0,47	1,20	52	104
15	canid (2,6%), capybara (2,0%), hare (5,2%), collared anteater (1,4%), armadillo (12,1%)	655,09	3,46	2.412,39	0,00	1,76	4,90	4,79	0,90	4,14	14,11	1,6	0,50	0,00	25	29
Average		**315,33**	**7,87**	**311,59**	**0,56**	**2,26**	**1,19**	**5,87**	**3,62**	**3,41**	**15,57**	**2,3**	**0,55**	**1,340**	**644**	**667**

was also run over in stretches with a high traffic volume, while lowland tapirs were run over in areas where traffic volume was low. Armadillos, South American raccoons, maned wolves, hares, deer, skunks, porcupines, and giant anteaters were run over predominantly at night, whereas capybaras, monkeys, lowland tapirs, and pumas were predominantly run over during the day.

CONCLUSION

The SOM analysis results revealed that the concentrations of wildlife-vehicle collisions were related to road characteristics and land cover. This allowed the relationships between multiple linear and non-linear variables, such as ecological data, to be analyzed.

The following interventions could be implemented in areas where road kill clusters occur. First, underpass tunnels for animals should be installed so that animals can cross the road safely and avoid collisions with vehicles moving at high speeds. Second, fences should be installed parallel to roads that are adjacent to wildlife crossings to funnel animals toward crossing structures (Van Der Ree 2015). Because wildlife-vehicle collisions occurred approximately 300 m from bodies of water and forest formations, the fences should have a minimum length of 300 m, on average, on either side of the wildlife crossing structure. Finally, tools such as mobile phone applications with which road users can report accidents in real-time (Tong et al. 2020) should be developed so that data from road users can be collected and a richer database on wildlife-vehicle collisions can be built.

Acknowledgments

The Authors thank Secretaria do Meio Ambiente do Estado de São Paulo (SMA) and the agencies Fundação Brasileira para o Desenvolvimento Sustentável (FBDS) and Instituto Nacional de Pesquisas Espaciais (INPE) for providing data. The Authors would also like to thank Companhia de Saneamento do Estado de São Paulo (CETESB) for providing the necessary data of collision to complete this study and to geologist Lucas Sandré by the final's figures and tables editions. The Author JAQ received financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (Process 305188/2020-8).

REFERENCES

ABRA FD. 2019. Mammal-vehicle collisions on toll roads in São Paulo State: implications for wildlife, human safety and costs for society. Doctoral dissertation, Universidade de São Paulo. (Unpublished).

ABRA FD, HUIJSER MP, MAGIOLI M, BOVO AAA & DE BARROS KMPM. 2021. An estimate of wild mammal roadkill in São Paulo state, Brazil. Heliyon 7(1): e06015.

AZEVEDO FC ET AL. 2013. Avaliação do risco de extinção da Onça-parda *Puma concolor* (*Linnaeus*, 1771) no Brasil. Biodivers Bras 3(1): 107-121.

AZEVEDO FC, LEMOS FG, FREITAS-JUNIOR MC, ARRAIS RC, MORATO RG & AZEVEDO FCC. 2021. The importance of forests for an apex predator: spatial ecology and habitat selection by pumas in an agroecosystem. Anim Conserv 24(3): 499-509.

BARTHELMESS EL. 2014. Spatial distribution of road-kills and factors influencing road mortality for mammals in Northern New York State. Biodivers Conserv 2(10): 2491-2514.

BENATTI D, DE SANTI M, WERTHER K, TEBALDI JH & HOPPE EGL. 2021. Helminthfauna of road-killed cougars (*Puma concolor*) from the Northeastern Region of São Paulo State, Brazil. Braz J Vet Parasitol 30(1): e024120. https://doi.org/10.1590/S1984-29612021008.

BÍL M, ANDRÁŠIK, R, SVOBODA, T & SEDONÍK J. 2016. The KDE+ software: a tool for effective identification and ranking of animal-vehicle collision hotspots along networks. Landsc Ecol 31(2): 231-237.

BUENO C, FAUSTINO MT & FREITAS SR. 2013. Influence of landscape characteristics on capybara road-kill on highway BR-040, southeastern Brazil. Oecologia Aust 17(2): 320-327.

CARNEIRO CDC, FRASER SJ, CRÓSTA AP, SILVA AM & BARROS CEDM. 2012. Semiautomated geologic mapping using self-organizing maps and airborne geophysics in the Brazilian Amazon. Geophysics 77(4): K17-K24.

CÉRÉGHINO R & PARK YS. 2009. Review of the Self-Organizing Map (SOM) approach in water resources: Commentary Environ Model Softw 24(8): 945-947.

CLEVENGER AP, CHRUSZCZ B & GUNSON KE. 2003. Spatial patterns and factors influencing small vertebrate fauna road-kill aggregations. Biol Conserv 109: 15-26.

COMPIN A & CÉRÉGHINO R. 2007. Spatial patterns of macroinvertebrate functional feeding groups in streams in relation to physical variables and land-cover in Southwestern France. Landsc Ecol 22(8): 1215-1225.

DAVIES DL & BOULDIN DW. 1979. A Cluster Separation Measure; IEEE Trans. Pattern Anal. Mach. Intell PAMI 1(2): 224-227.

DE FREITAS SR, DE OLIVEIRA NA & CIOCHETI G. 2015. How landscape features influence road-kill of three species of mammals in the brazilian savanna? Oecologia Aust 18: 35-45.

DE OLIVEIRA TG ET AL. 2010. Ocelot ecology and its effect on the small-felid guild in the lowland neotropics. In: Biology and Conservation of Wild Felids. New York, NY: Oxford University Press Inc, p. 559-580.

FBDS - FUNDAÇÃO BRASILEIRA PARA O DESENVOLVIMENTO SUSTENTÁVEL. 2017. Mapeamento do uso do solo dos biomas Cerrado e Mata Atlântica na escala 1:20.000, Rio de Janeiro.

FORMAN RTT. 1998. Road ecology: a solution for the giant embracing us. Landsc Ecol 13(4).

FREITAS SR, HAWBAKER TJ & METZGER JP. 2010. Effects of roads, topography, and land cover on forest cover dynamics in the Brazilian Atlantic Forest. For Ecol Manage 259(3): 410-417.

GLISTA DJ, DEVAULT TL & DEWOODY JA. 2008. Vertebrate road mortality predominantly impacts amphibians. Herpetol Conserv Biol 3(1): 77-87.

GLISTA DJ, DEVAULT TL & DEWOODY JA. 2009. A review of mitigation measures for reducing wildlife mortality on roadways. Landsc Urban Plan 91(1): 1-7.

GUNSON KE, MOUNTRAKIS G & QUACKENBUSH LJ. 2011. Spatial wildlife-vehicle crash models: A review of current work and its application to transportation mitigation project. J Environ Manage 92(4): 1074-1082.

HSU CC. 2006. Generalizing self-organizing map for categorical data. IEEE Trans Neural Networks 17(2): 294-304.

HUIJSER MP, ABRA FD & DUFFIELD JW. 2013. Mammal road mortality and cost-benefit analyses of mitigation measures aimed at reducing crashes with capybara (Hydrochoerus hydrochaeris) in São Paulo State, Brazil. Oecologia Aust 17(1): 129-146.

HUIJSER MP & BERGERS PJM. 2000. The effect of roads and traffic on hedgehog (Erinaceus europaeus) populations. Biol Conserv 95: 111-116.

HUIJSER MP, DUFFIELD JW, CLEVENGER AP, AMENT RJ & MCGOWEN PT. 2009. Cost-benefit analysis of mitigation measures aimed at reducting crashes with large ungulates in the United States and Canada, a decision support tool. Ecol. Soc 14(2): 15.

KOHONEN T. 2013. Essentials of the self-organizing map. Neural Networks 37: 52-65.

KUSSUL N, LAVRENIUK M, SKAKUN S & SHELESTOV. 2017. A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. IEEE Geosci Remote Sens Lett 14(5): 778-782.

LITVAITIS JA & TASH JP. 2008. An approach toward understanding wildlife-vehicle collisions. Environ Manage 42(4): 688-697.

MORELLE K, LEHAIRE F & LEJEUNE P. 2013. Spatio-temporal patterns of wildlife-vehicle collisions in a region with a high-density road network. Nature Conserv (5): 53-73.

PAGANY R. 2020. Wildlife-vehicle collisions-Influencing factors, data collection and research methods. Biol Conserv 251: 108758.

PARK YS, CÉRÉGHINO R, COMPIN A & LEK S. 2003. Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. Ecol Modell 160(3): 265-280.

ROSS JLS & MOROZ IC. 1997. Geomorphological map of São Paulo State. São Paulo: USP/FAPESP/IPT.

ROWDEN P, STEINHARDT D & SHEEHAN M. 2008. Road crashes involving animals in Australia. Accid Anal Prev 40(6): 1865-1871.

SÁENZ-DE-SANTA-MARÍA A & TELLERÍA JL. 2015. Wildlife-vehicle collisions in Spain. Eur J Wildl Res 61(3): 399-406.

SCHWARTZ AL, SHILLING FM & PERKINS SE. 2020. The value of monitoring wildlife roadkill. Eur J Wildl Res 66(1): 1-12.

SHILLING F, COLLINSON W, BIL M, VERCAYIE D, HEIGL F, PERKINS SE & MACDOUGALL S. 2020. Designing wildlife-vehicle conflict observation systems to inform ecology and transportation studies. Biol Conserv 251: 108797.

SPELLERBERG IF & FEDOR PJ. 2003. A tribute to Claude Shannon (1916 – 2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon-Wiener' Index Glob Ecol Biogeogr 12(3): 177-179.

TAYLOR SK, BUERGELT CD, ROELKE-PARKER ME, HOMER BL & ROTSTEIN DS. 2002. Causes of Mortality of Free-Ranging Florida Panthers. J Wildl Dis 38(1): 107-114.

TEIXEIRA FZ ET AL. 2013. Are road-kill hotspots coincident among different vertebrate groups?. Oecologia Aust 17(1): 36-47.

TONG Y, ZHOU Z, ZENG Y, CHEN L & SHAHABI C. 2020. Spatial crowdsourcing: a survey. VLDB J 29(1): 217-250.

TREVELIN LC, PORT-CARVALHO M, SILVEIRA M & MORELL E. 2007. Abundance, habitat use and diet of *Callicebus nigrifrons Spix* (*Primates, Pitheciidae*) in Cantareira State Park, São Paulo, Brazil. Rev Bras Zool 24: 1071-1077.

VALERIO F, BASILE M & BALESTRIERI R. 2021. The identification of wildlife-vehicle collision hotspots: Citizen science reveals spatial and temporal patterns. Ecol Process 10(1): 1-13.

VAN DER REE R, GAGNON JW & SMITH DJ. 2015. Fencing: A Valuable Tool for Reducing Wildlife-Vehicle Crashes and Funnelling Fauna to Crossing Structures. Handb Road Ecol 159-171.

VESANTO J, HIMBERG J, ALHONIEMI E & PARHANKANGAS J. 2000. SOM Toolbox for Matlab 5.

How to cite

TSUDA LS, CARNEIRO CC & QUINTANILHA JA. 2022. Using Self-Organizing Maps to find spatial relationships between wildlife-vehicle crashes and land use classes. An Acad Bras Cienc 94: e20210727. DOI 10.1590/0001-3765202220210727.

Manuscript received on May 19, 2021;
accepted for publication on September 10, 2021

LARISSA S. TSUDA[1]
https://orcid.org/0000-0002-4468-1499

CLEYTON C. CARNEIRO[2]
https://orcid.org/0000-0002-4032-200X

JOSÉ ALBERTO QUINTANILHA[3]
https://orcid.org/0000-0003-3261-7825

[1]Programa de Pós Graduação em Engenharia de Transportes (PPGET), Escola Politécnica da Universidade de São Paulo (EPUSP), Departamento de Engenharia de Transportes, Av. Professor Almeida Prado, Trav. 2, 83, 05508-900 São Paulo, SP, Brazil

[2]Escola Politécnica, Universidade de São Paulo, Departamento de Engenharia de Minas e de Petróleo, Campos Santos, Praça Coronel Narciso de Andrade, s/n, Vila Mathias, 11013-560 Santos, SP, Brazil

[3]Universidade de São Paulo, Instituto de Energia e Ambiente/IEE, Divisão Científica de Gestão, Ciência e Tecnologia Ambiental, Av. Prof. Luciano Gualberto, 1289, 05508-900 São Paulo, SP, Brazil

Correspondence to: **José Alberto Quintanilha**
E-mail: jaquinta@usp.br

Authors contributions

Larissa Sayuri Tsuda contributes to Data collection; Data analysis and interpretation; Draft the article; Critical revision of the article. Cleyton de Carvalho Carneiro contributes to the Design of the work; Data analysis and interpretation; Draft the article; Critical revision of the article. José Alberto Quintanilha contributes to the Concept of the work; Data analysis and interpretation; Draft the article; Critical revision of the article; Final approval of the version to be published.