European Journal of Geography Volume 10, Number 3:137-153, September 2019

USING GEOSPATIAL ANALYSIS TECHNIQUES FOR EVALUATING THE ASSOCIATION BETWEEN SOCIOENVIRONMENTAL FACTORS AND THE GEOGRAPHICAL DISTRIBUTION OF LEPTOSPIROSIS IN SAO PAULO, BRAZIL

Marcos FERREIRA
Campinas State University - UNICAMP, Geosciences Institute, Department of Geography
macferre@ige.unicamp.br

Marta MARUJO FERREIRA
Unifal-MG, Department of Geography
martafelicia@uol.com.br

Abstract

Leptospirosis is a disease that is caused by the *Leptospira* bacteria. Transmission occurs via contact with water and mud that has been contaminated by the rodent's urine. High incidence rates occur during wet seasons. The aim of this research was to use geospatial techniques to evaluate the association between the socioenvironmental variables and the spatial distribution of leptospirosis in Sao Paulo, Brazil. Four physical geographic maps (slope, elevation, distance from rivers and river density) and two socioeconomic maps (average family salary and average number of residents in the household) were used. The statistical significance of the differences between the average values of the variables that are calculated for the areas with leptospirosis cases and the areas without leptospirosis cases was evaluated. The average family salary, the number of residents in the households and the river density from the census sector variables were significantly associated with cases of leptospirosis.

Keywords: *Spatial analysis, GIS, Leptospirosis, Sao Paulo, Health Geography, Brazil*

1. INTRODUCTION

Recent studies have focused on modeling epidemics using geospatial analysis, including dengue fever (Ferreira, 2014; Mahmood et al., 2019); yellow fever (Kraemer et al., 2017), A(H1N1)pdm09 (Doukissas et al., 2018); and leptospirosis (Luenam and Puttanapong, 2019). Leptospirosis is a disease with a worldwide geographical distribution. It is caused by bacteria in the Leptospira genus, and it affects humans and animals. Transmission to humans occurs directly by contact with rodent urine or tissue and indirectly by contact with water or mud that have been contaminated by rodent urine (Caldas, 1979; Levett, 2001; Raghavan et al., 2012, 2013). Rodents, mainly the *Rattus norvegicus*, are the hosts of the bacteria. The disease affects people working with animals in open spaces or living in urine-contaminated environments (Traxler et al., 2014; Hartskeerl et al., 2011).

In tropical regions, higher incidence rates are frequent during wet seasons or after major river floods (Ward et al., 2004; Gracie et al., 2014). Several urban geographical factors are associated with the bacteria infection, such as family income, river proximity, sanitary conditions, land use, terrain slope and household characteristics (Barcelos and Sabroza, 2001; Reis, 2008; Soares et al., 2010; Robertson et al., 2012; Raghavan et al., 2012; Gracie et al., 2014; Vega-Corredor and Opadeyi, 2014). These factors interact spatially, creating geographical local associations that favour bacteria dissemination throughout the population. In the large cities of developing countries, leptospirosis has been associated with living near flood areas, near domestic garbage disposal sites and in densely populated slums near river channels and periodically flooded areas (Reis, 2008). In Brazil, leptospirosis epidemics are concentrated from the months of October-March, when major floods may distribute the Leptospira bacteria throughout populations living near rivers.

Our study proposes an exploratory spatial analysis of leptospirosis using GIS techniques. It evaluates the effects of the socioenvironmental variables on the spatial distribution of leptospirosis cases in Sao Paulo - the most populous municipality in Brazil with 12,100,000 inhabitants (SEADE, 2018). Figure 1 shows the location of the Sao Paulo municipality in Brazil.

The study analysed four physical geographic risk variables (slope, elevation, distance from rivers and river density) and two socioeconomic risk variables (average family salary and number of residents in the household). Slope and elevation are geomorphological variables that directly influence the flooding process along river plains and the surficial water accumulation sites that are created in urban areas. We used Quantum GIS Desktop GIS 3.02 (QGIS) to perform the geospatial analysis operations, data generation and thematic mapping of the environmental and socioeconomic variables.

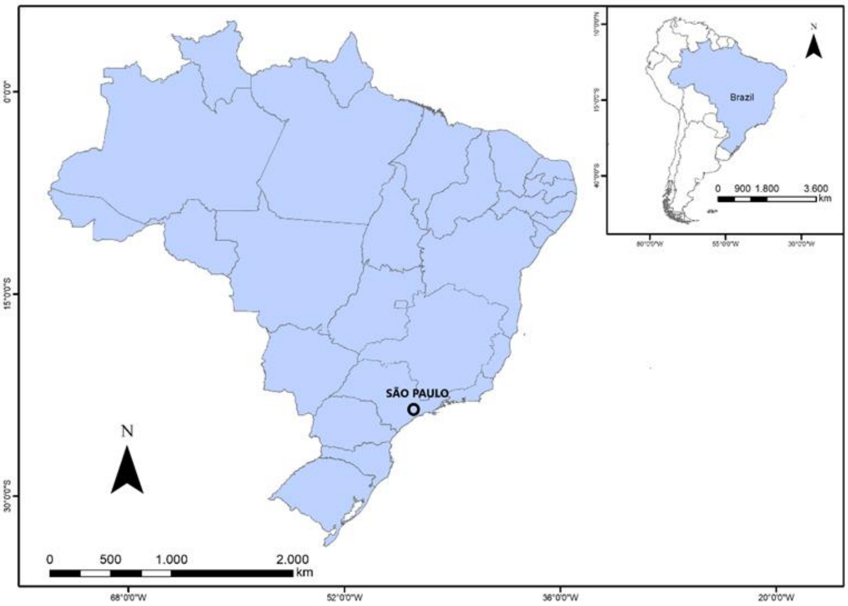

Figure 1. Location map of the Sao Paulo municipality in Brazil.

2. LITERATURE REVIEW

The first study about leptospirosis in Brazil, which was carried out by Azevedo and Correa (1968) in Recife, concluded that the disease's epidemics occurred after major flood events. A few years later, in Rio de Janeiro, Silva et al. (1975) noted that there was an association between the number of infected people and the hydrological and socioeconomic characteristics of surrounding areas. Barcelos and Sabroza (2001) concluded that infection risk is lower farther from water sites and higher closer to the exposed garbage areas of Rio de Janeiro. Leptospirosis cases are more frequent in Rio de Janeiro during the summer when populations have direct contact with river and lake waters in response to high temperatures (Guimaraes et al., 2014).

In Belo Horizonte, Brazil, it was found that 24% of the infected people lived in slums and poor neighbourhoods, and 21% lived in lowlands that were exposed to periodic flooding (Figueredo et al., 2001). In Sao Paulo, during the dry season, high incidence rates occurred in low-income neighbourhoods that were distant from rivers; however, during the wet season, the disease also spreads to poor neighbourhoods that were situated near rivers (Soares et al., 2010).

Leptospirosis transmission occurs mainly in densely populated suburbs that are a short distance from rivers in Sri Lanka (Robertson et al., 2012). In Semarang city, Java, the cases are concentrated in the dry season and in areas that are more elevated and with no flooding risk, which is different from those in other tropical countries (Sunaryo, 2012). A study that was published by Sanchez-Montes (2015) shows that in Mexico, air temperature is more important in explaining the spatial distribution of leptospirosis than rainfall. In Lyon, France, the transmission rate is large in densely populated and low-income neighbourhoods. In these areas, more rodents that were infected with Leptospira bacteria were found (Ayral et al., 2015).

Furthermore, the association of the environmental and socioeconomic factors with the locations of leptospirosis cases is affected by the geographical scale and the areal unit size that is used in the geospatial analysis. A study that was carried out in the state of Rio de Janeiro, Brazil, Gracie et al. (2014) concluded that at the local level (census sector), leptospirosis incidence was correlated with the percentage areas that flooded; at the regional scale, the incidence was correlated with the number of people living in slums and the percentage of densely urbanized areas. However, at the municipal scale, the authors observed that there were no significant correlations between the environmental and socioeconomic factors and leptospirosis incidence.

An extensive study carried out by Soares et al. (2010) evaluated the association between socioeconomic variables and leptospirosis incidence and lethality in Sao Paulo using a district database from 1998 to 2006. The results showed that the spatial pattern of clustered cases was related to the literacy rate, average monthly income, number of residents per household, water supply and sewage system. These authors also found that the incidence and lethality rates were correlated with the socioeconomic conditions of the population in both the rainy and dry seasons. These authors also found that during the dry season, the cases were concentrated only in the poorest districts of the city, but during the rainy season, the disease also spread to the other districts, possibly due to the districts’ proximity to rivers. The research of Soares et al. (2010) did not evaluate the effects of the environmental variables on the distribution of leptospirosis cases, although it has been suggested that their influence is probably greater during the rainy season.

Rapid changes in the urban landscape structure of large cities may be affecting the spatial epidemiology of vector-borne infectious diseases in some Latin American countries, such as Brazil. Those changes create new epidemiological landscape units that are composed of areas where the host, vector and pathogen interact spatially within an environment that is permissive to transmission (Reisen, 2010). The most significant human-induced impact on the urban landscape is the creation of domestic microhabitats (nidus) that favour the transmission of diseases (Pavloskiy, 1966; Gubler, 1996). Leptospirosis microhabitats are formed by the spatial association between environmental factors (river flooding, flat terrain, areas that are located near urban rivers, high river density and low elevation areas), and socioeconomic factors (poverty, average monthly salary, and number of residents per household, among others).

3. MATERIALS AND METHODS

3.1 Materials

3.1.1 Vector cartographical database

A cartographic database that was composed of four vector maps in the shapefile format was used as follows. The maps included the territorial limits of the Sao Paulo Municipality, the hydrographical network, the census sectors and the districts of Sao Paulo. All these shapefiles were obtained from the Metropolis Study Centre of Sao Paulo (CEM, 2015; SIH, 2007) and GEOSAMPA (2017) and later projected on a South American Datum 1969 - SAD 69, and UTM coordinate system using a QGIS projection module. The census sector is the smallest spatial aggregation unit that is used by the Brazilian Institute of Geography and Statistics (IBGE, 2010) to collect socioeconomic information in the national census survey. From the census sector shapefiles, data for the average salary, total resident population and average number of people living in the household were gathered.

3.1.2 Leptospirosis data

A total of 1,885 cases that were documented over 10 years (1998-2007), 1,702 of which were from 1998-2006 and 183 from 2007, were used. The 1998-2006 data were obtained from a leptospirosis case map that was available in the Identification and Delimitation of Priority Areas for Leptospirosis Control in Sao Paulo Report (PMSP, 2007). This map, which was originally in raster format, was georeferenced in QGIS using an SAD69 datum and latitude-longitude coordinate system to represent the locations of all cases that occurred over the 1998-2006 period. A sample of 1,702 leptospirosis cases from the period from 1998-2006 was captured using an on-screen digitizing method that was applied to the leptospirosis case map at a 1:50,000 scale by means of the Vector Point Feature Adding option in the Editing module of QGIS. Only the visually accurate points in the leptospirosis case raster map at that scale were digitized.

In addition to the 1,702 leptospirosis cases that were digitized and collected for the period from 1998-2006, data from 2007 that were available in point shapefile vector format, containing the locations of 183 geocoded leptospirosis cases, were obtained from the Metropolis Studies Centre of Sao Paulo database (SIH, 2007; CEM, 2014); they were added to the database, increasing the total to 1,885 cases.

3.1.3 Socioenvironmental variables data

- *Slope (SLO) and surface elevation (ELV):*Data from ASTER GDEM2 (Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model) were used to map the slope and surface elevation of the Sao Paulo Municipality. ASTER GDEM2 data represent the terrain elevations using a 30-metre spatial resolution grid and a 15-metre altimetry resolution (Tachikawa et al., 2011). The digital elevation model data in raster format were obtained from ASTER GDEM (2014). The SLO and ELV maps were generated using the Terrain Analysis Raster module of QGIS.
- *Average monthly family salary (SAL) and Average number of residents in the household (RES):* Data from the SAL and RES variables were gathered from the census sector database (IBGE, 2010). This database includes a census sector map and an attribute table containing the demographic and socioeconomic data that are linked to census sector polygons. The SAL and RES maps were prepared using the field calculator tool that is available in the Vector Attribute Table module. Since the RES variable refers to the average number of residents in the household, in sparsely populated areas, uninhabited households may exist, and for this reason, the average value of the RES variable can be less than 1.0.
- *River distance (RDI):*The spatial variation of distances from rivers was represented as an isodistance raster map, which was prepared using a 30-metre spatial resolution grid. The grid cells of the map include the recorded Euclidian distances from a certain point to the nearest river. The river distances were mapped using a hydrographic network map

as a reference and the Distance function, which was available in the Raster module of QGIS.

- *River density in the census sector (RDE):*River density in the census sector was calculated using hydrographic and census sector maps. These maps were combined by using the Adding Line Lengths in Polygons function, which was available in the Vector module of QGIS. Then, the total river channel length by census sector polygons was extracted. Next, the river density in the census sector was calculated by using the Field Calculator tool, which was available in the Vector Attribute Table module of QGIS.

3.2 Methods

3.2.1 Leptospirosis average annual incidence rate by Municipal District

In the first step of the research, we calculated the leptospirosis incidence rate in the 96 districts of Sao Paulo. In this sense, the average annual incidence rate per 100,000 inhabitants (I_L) in the period from 1998-2007 was estimated using Equation 1:

$$I_L = \left(\frac{n}{P_{avg}}\right).100{,}000 \qquad \text{(Eq. 1)}$$

where P_{avg} is the total average population in the districts in 2000 and 2010, and n is the total number of reported leptospirosis cases by district in the 1998-2007 period. The I_L values were mapped into four classes using the quantile classification method. A population density map by district was also produced to show if the districts with the highest population densities are those with the highest incidence.

3.2.2 Kernel density estimation (KDE)

The spatial density of leptospirosis cases was mapped using the kernel density estimation (KDE) method, which uses a mathematical smoothing function for the aggregate point nuclei mapping that is distributed on a grid surface (Smith et al, 2009). We used the quartic bi-weight function and a 3,300-metre radius around the leptospirosis case points as the spatial parameters of the kernel method. The choice of the radius value was based on the maximum daily displacement of rodents (3,300 m), which was estimated by Fenn and MacDonald (Fenn and MacDonald, 1987; Masi et al., 2009) using telemetric techniques. The kernel density map shows the possible rodent activity areas that are located around areas with a high density of leptospirosis point cases. The KDE map was classified into four classes, varying from a low density to a very high density of leptospirosis cases, using the quantile method.

The kernel density values that were classified as very high kernel density (VHD) were considered to be the areas with a great number of microhabitats for leptospirosis. Therefore, the VHD areas are examples of the possible epidemiological landscape units of the disease. The land use spatial structure of the VHD areas was visualized and analysed using high-resolution orthorectified digital aerial photographs (1-metre spatial resolution).

3.2.3 Spatial autocorrelation analysis

The presence of spatial autocorrelation (SA) for the I_L variable was also tested using the univariate LISA (local univariate spatial autocorrelation) with Moran's index (I). The K-nearest neighbours spatial weight option, which was available in the GeoDa package (Anselin, 2003), was used to calculate the I values. We chose Moran's index because it is one of the most commonly used indices in spatial analysis and is easily accessible in the GeoDa software. The results are presented as an SA scatter plot and a positive SA district map that show the districts with a significant local Moran's index.

3.2.4 Statistical Analysis

A grid with cells measuring 200 x 200 metres was overlaid on the map of Sao Paulo. This grid was used in the National Census of Brazil 2010 survey. From this grid, we selected 1,633 cells where cases of leptospirosis were registered during the period from 1998 to 2007. The selection of the cells was made using the Point in the Polygon vector operation and the Attribute Table feature selection option that were available on QGIS. In 1,590 cells (97.36%), one case occurred, and in the other 43 cells (2.63%), two cases of the disease occurred. The map of these cells, which are called case cells, shows the samples from areas possibly containing microhabitats of leptospirosis. The population residing in the cells was obtained from the Brazilian National Census of 2010.

Descriptive statistical parameters of the resident populations in the case cells were calculated, and the outliers were removed. In this way, only a population ranging from 306 to 1,001 people residing in the case cells was considered. Based on this population interval, we selected a sample of 1,028 case cells (Figure 2a). In addition, a no-case cell map was produced by selecting only the cells whose resident populations were in the same population interval of the case cells [306 – 1,001]. Thus, a sample of 653 cells in which no cases were recorded was selected (Figure 2b).

The average values of the SAL, RES, RDI, RDE, SLO and ELV variables were calculated for the case cell and no-case cell maps using the Statistical by Zone operation, which is available in the Raster module of the QGIS. Since the SAL, RES and RDE maps were in vector polygonal format (census sectors), their conversion to the raster format was necessary.

First, the centroids of the census sectors were mapped, and the values of the SAL, RES and RDE variables were assigned to these centroids. Then, the centroid values were interpolated by using the inverse square distance algorithm using the same spatial resolution as the other maps (30 m). Then, the independent samples t-test was chosen to evaluate the statistical significance of the differences between the average values of the SAL, RES, RDI, RDE, SLO and ELV variables, which were calculated for the case cell and for the no-case cell maps. The statistical analysis was performed using the *MedCalc* statistical software version 18.6 (MedCalc, 2018).

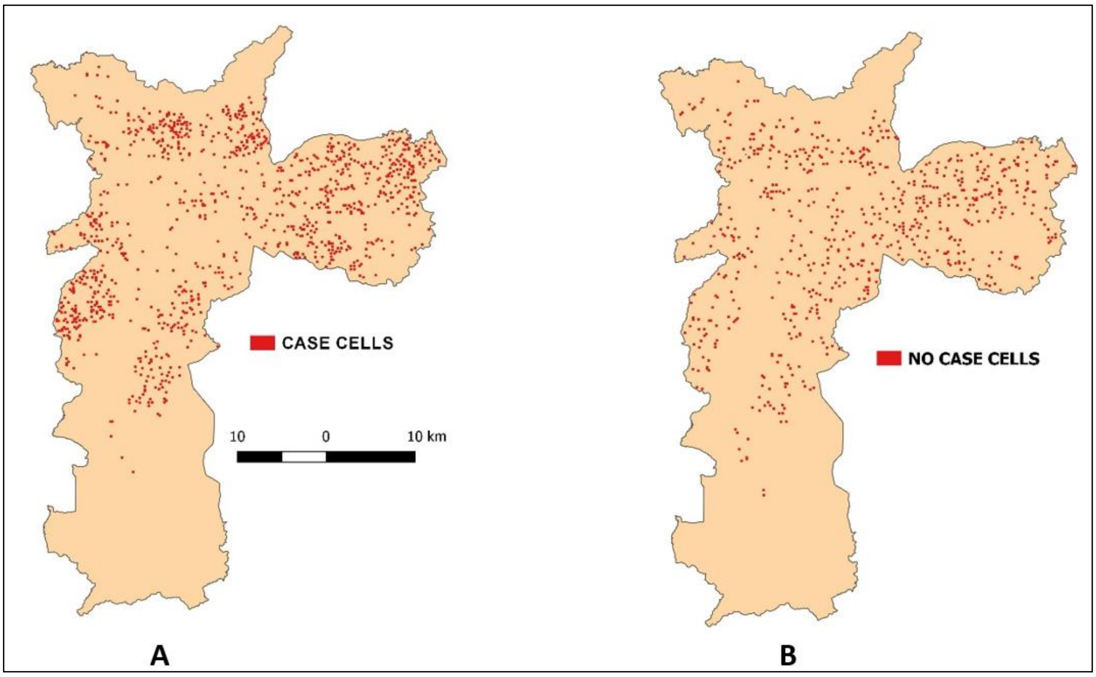

Figure 2. The case cell (A) and no-case cell (B) maps that were used as a reference to evaluate the statistical significance of the differences between the average values of the SAL, RES, RDI, RDE, SLO and ELV variables.

4. RESULTS

4.1 Average Leptospirosis Incidence Rate by Municipal District

The average annual incidence rate (I_L) of leptospirosis in the 1998-2007 period in Sao Paulo was 2.048 cases per 100,000 (95% CI = 1.781 ± 0.1770). Districts with higher I_L values, which were classified in the upper quartile, are shown in Table 1. Figure 3a shows the I_L values by district map. It can be noted on the I_L map that the disease is concentrated predominantly in districts that are located in the outlying areas of the municipality.

We could identify two large clusters of districts with higher incidences: the Southwest region, which formed by the districts of Socorro, Santo Amaro, Capao Redondo, Campo Limpo, Morumbi, Vila Sônia, Raposo Tavares, Rio Pequeno, Butanta, Jaguare and Jaguara; and the Northeast region, which is formed by the districts Tremembe, Jaçana, Vila Medeiros, Vila Guilherme and Cachoeirinha. It was also possible to identify other districts with high incidences that are distributed in the eastern zone and in the central zone, where the Se district is located, and with the highest estimated incidence of the disease (I_L = 4.318)

The KDE map (Figure 3b) shows the spatial distribution of five areas with a very high density of cases (VHD) that were accumulated over 10 years. VHD 1 includes the Cachoerinha and Brasilândia districts; VHD 2 includes the Tremembe and Jaçana districts; VHD 3 includes the Vila Curuça and Itaim Paulista districts; VHD 4 includes the Cidade Ademar district; and VHD 5 includes the Capao Redondo, Campo Limpo and Jardim Sao Luiz districts. By comparing the KDE map with the population density map (Figure 3c), we can see that the very high density of cases areas 3, 4 and 5 are related to the high population density districts, such as Vila Curuça (east zone), Cidade Ademar (southeast zone) and Capao Redondo (southwest zone).

Table 1. Leptospirosis average incidence rate (L_I) (1998-2007) by municipal district in Sao Paulo (districts classified in the L_I upper quartile). The geographical location of districts, which are directly identified using ID numbers, is shown in Figure 3b. The L_I values were calculated using Equation 1.

ID	District	Population	n	L_I 95% CI = 1.781 ± 0.1770
1	Se	20,840	9	4.318
2	Parque do Carmo	66,849	27	4.038
3	Cachoeirinha	152,529	59	3.868
4	Morumbi	33,435	12	3.589
5	Jaçana	92,323	33	3.574
6	Bras	25,873	9	3.478
7	Vila Medeiros	134,628	45	3.342
8	Raposo Tavares	94,059	30	3.189
9	Barra Funda	12,971	4	3.083
10	Butanta	50,345	15	2.979
11	Erm. Matarazzo	111,735	33	2.953
12	Tremembe	174,767	50	2.860
13	Jaguare	42,054	12	2.853
14	Vila Guilherme	49,196	14	2.845
15	Campo Limpo	203,813	57	2.796
16	Socorro	37,894	10	2.638
17	Capao Redondo	258,012	68	2.635
18	Rio Pequeno	113,878	30	2.634
19	Vila Curuça	146,482	40	2.589
20	Vila Sônia	87,379	22	2.481

21	Santo Amaro	60,539	15	2.481
22	Jaguara	24,914	6	2.408
23	Jardim Helena	146,370	35	2.391
24	Cidade Líder	123,548	28	2.266

4.2 - Spatial Autocorrelation

The univariate LISA map for the L_I variable is presented in Figure 3d. Moran's index for the leptospirosis incidence rate was low for a positive SA (Moran's $I = 0.2729$) for the 1998-2007 period. By analysing the map in Figure 3d, we noted that of the 30 positive SA districts (5.20% of all Sao Paulo districts), the SA *p-value* was 0.05 in five of them: Tucuruvi, Bras, Rio Pequeno, Vila Sônia, and Vila Andrade e Campo Limpo. The remaining positive SA districts had less significant I-values ($p > 0.05$).

The region with the greatest number of positive SA values was composed of the following contiguous districts: Jaguara, Jaguare, Rio Pequeno, Butanta, Raposo Tavares, Vila Sônia, Morumbi, Campo Limpo, Vila Andrade, Jardim Ângela, Socorro, Campo Grande and Pedreira.

We also identified three smaller positive SA clusters: North (Tremembe, Jaçana, Vila Medeiros, Tucuruvi and Brasilândia), East (Cidade Lider, Parque do Carmo, Jose Bonifacio, Sao Mateus and Iguatemi, Vila Curuça and Jardim Helena) and Central (Bras, Pari and Bom Retiro).

Figure 4 presents the orthorectified aerial photographs (1-metre spatial resolution) showing part of VHD areas 1 to 4, which were mapped in Figure 3b. When analysing these images, we see that the spatial structure of the urban landscapes in these places is characterized by the high land occupation density of houses that were constructed in a spontaneous way with very small lots. There are several lots with multiple houses that are located near rivers where untreated sewage is dumped, on rainwater accumulation sites, or on high-inclination angle hills. According to the Sao Paulo City Hall (PMSP, 2017), in these areas, there are many slums and precarious allotments where citizens live with high and very high social vulnerability.

The geotechnical map of the Sao Paulo Municipality (GEOSAMPA, 2017) shows the occurrence of the sites with accumulated garbage on the surface (*lixões*, in Portuguese) near these areas. These various socioenvironmental factors that are associated with these locations contribute to the formation of a landscape structure that is highly favourable to the emergence of leptospirosis. The spatial distribution of the leptospirosis cases in relation to the socioeconomic and physical geographic variable maps are shown in Figure 5.

4.3 - Statistical Analysis

The results of the statistical analysis are shown in Table 2 and Figure 6. In relation to the topographical variables, we found that the average slope in the case cells (SLO = 7.111°) was higher than that in the no-case cells (SLO=6.928°) ($p=0.1200$). The average altitude was higher in the case cells (ELV=775.07 m) than that in the no-case cells (ELV=774.86 m) ($p=0.8679$). In relation to the hydrographic variables, we observed that the distance to rivers in the case cells (RDI=141.80 m) was lower than that in the no-case cells (RDI=152.73 m) ($p=0.1025$). Furthermore, the river density in the census sector in the case cells (RDE=2.547 km/km^2) was higher than that in the no-case cells (RDE=2.309 km/km^2) ($p=0.0012$).

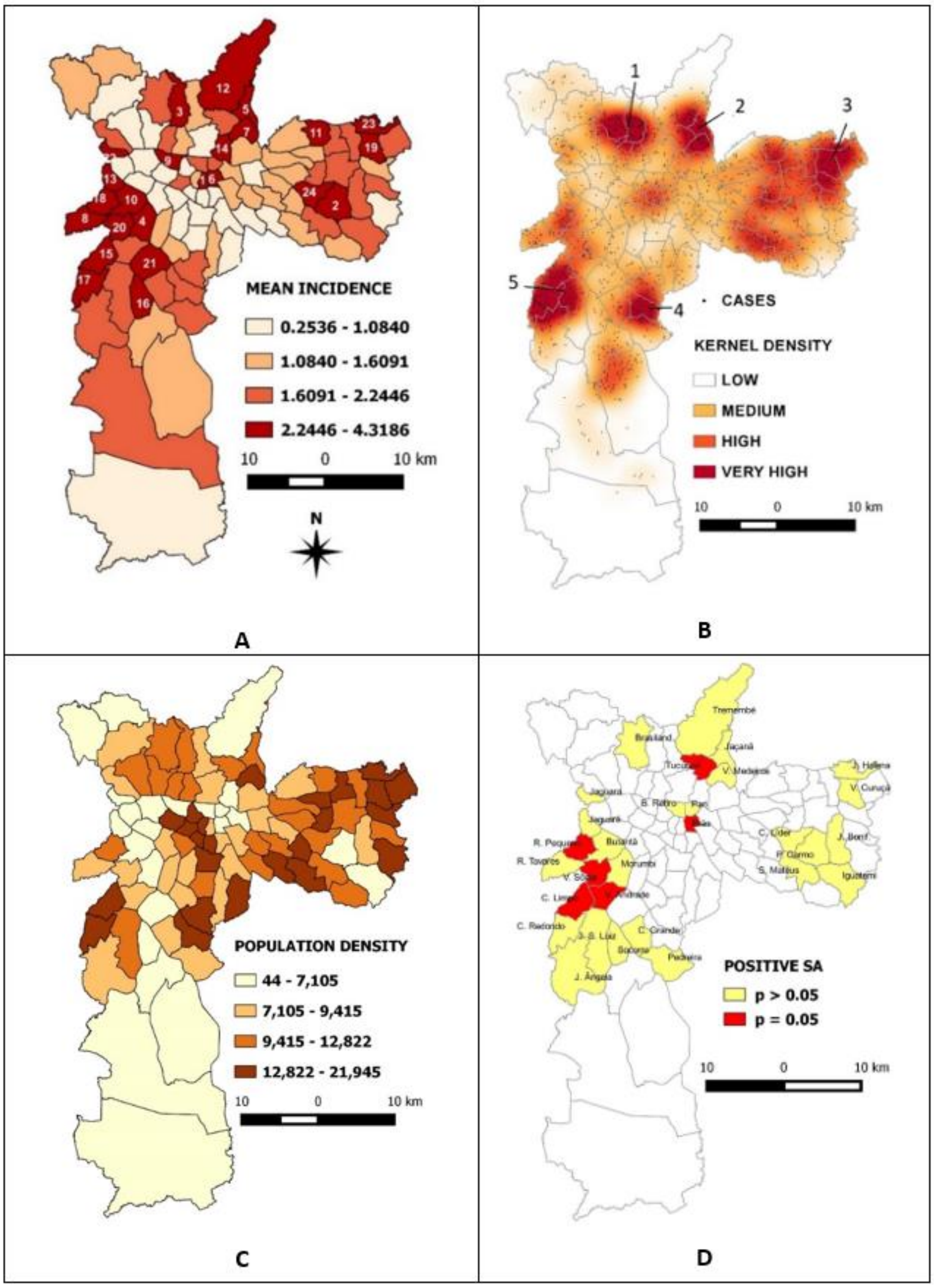

Figure 3. (A) Average leptospirosis incidence rate by district map, (B) the kernel density map showing in numbers the areas with a very high kernel density (VHD) of leptospirosis cases, the average population density by district map (C), and the LISA significance by district map showing the positive spatial autocorrelation districts for the leptospirosis incidence rate (D).

Table 2. Average values for the SAL, SLO, RDE, RDI, RES and ELV variables that were calculated for the case cell and no-case cell maps and their respective p-values that were estimated using the independent samples t-test.

	Case cells (n = 1,027)		No-case cells (n = 653)		
Variables	Average	95% CI	Average	95% CI	*p*
SAL (US$)	386.14	371.75 to 400.53	479.52	455.17 to 503.86	*<0.0001*$^{(1)}$
RES	3.195	3.180 to 3.210	3.091	3.067 to 3.115	*<0.0001*$^{(1)}$
SLO (º)	7.111	6.961 to 7.260	6.928	6.753 to 7.103	*0.1200*

RDE (km/km^2)	2.547	2.456 to 2.638	2.309	2.199 to 2.419	*0.0012*[(1)]
RDI (m)	141.80	133.63 to 149.97	152.73	142.44 to 163.022	*0.1025*
ELV (m)	775.07	773.53 to 776.60	774.86	772.86 to 776.85	*0.8679*

[(1)] *Differences between the average values that were statistically significant at $p<0.005$.*

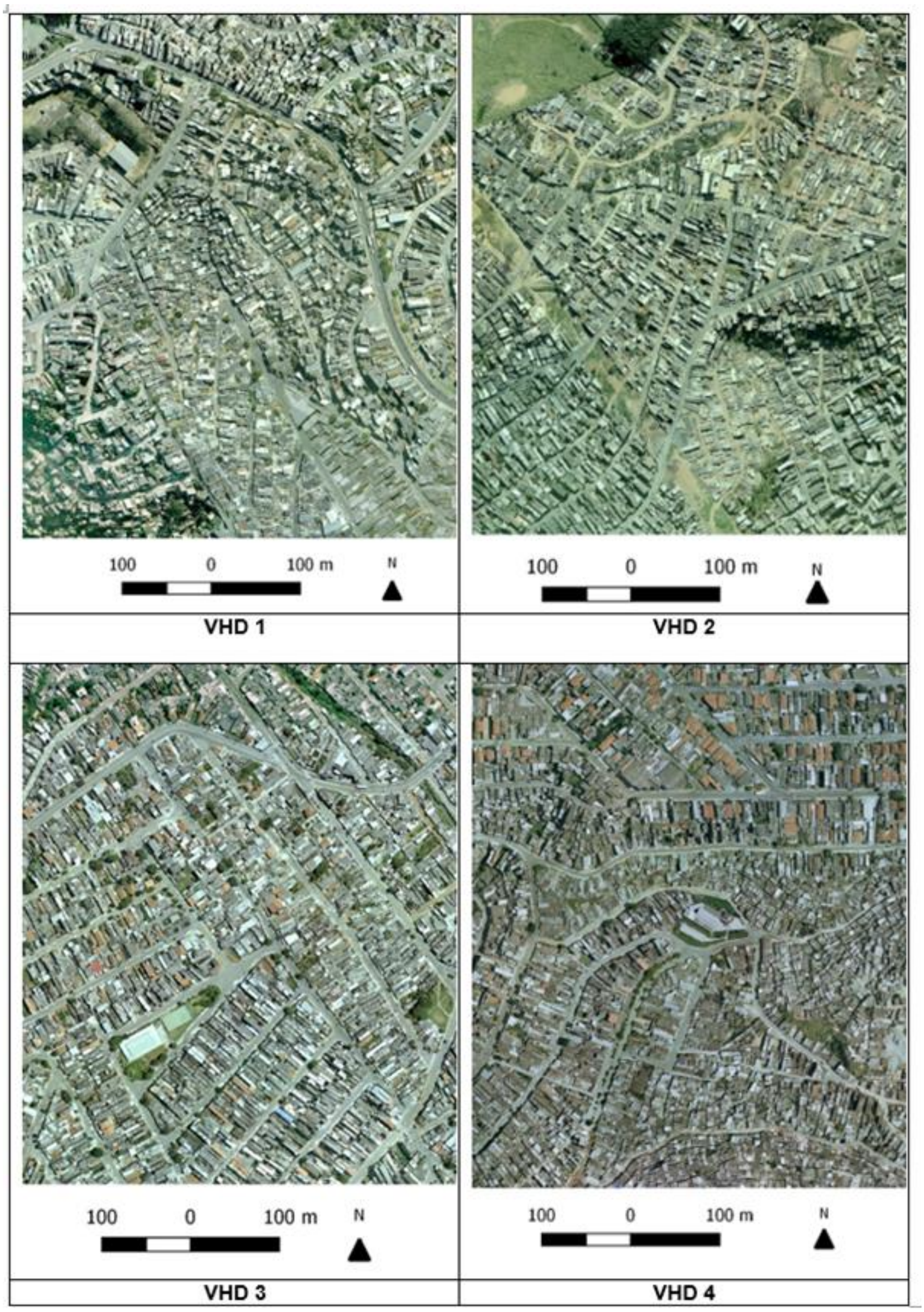

Figure 4. Partial view of the land use patterns in the very high kernel density of leptospirosis case areas (VHDs 1 to 4; see Figure 3b), which were obtained from high-resolution orthorectified aerial photos. Source: Aerial photos that were adapted from GEOSAMPA (2017).

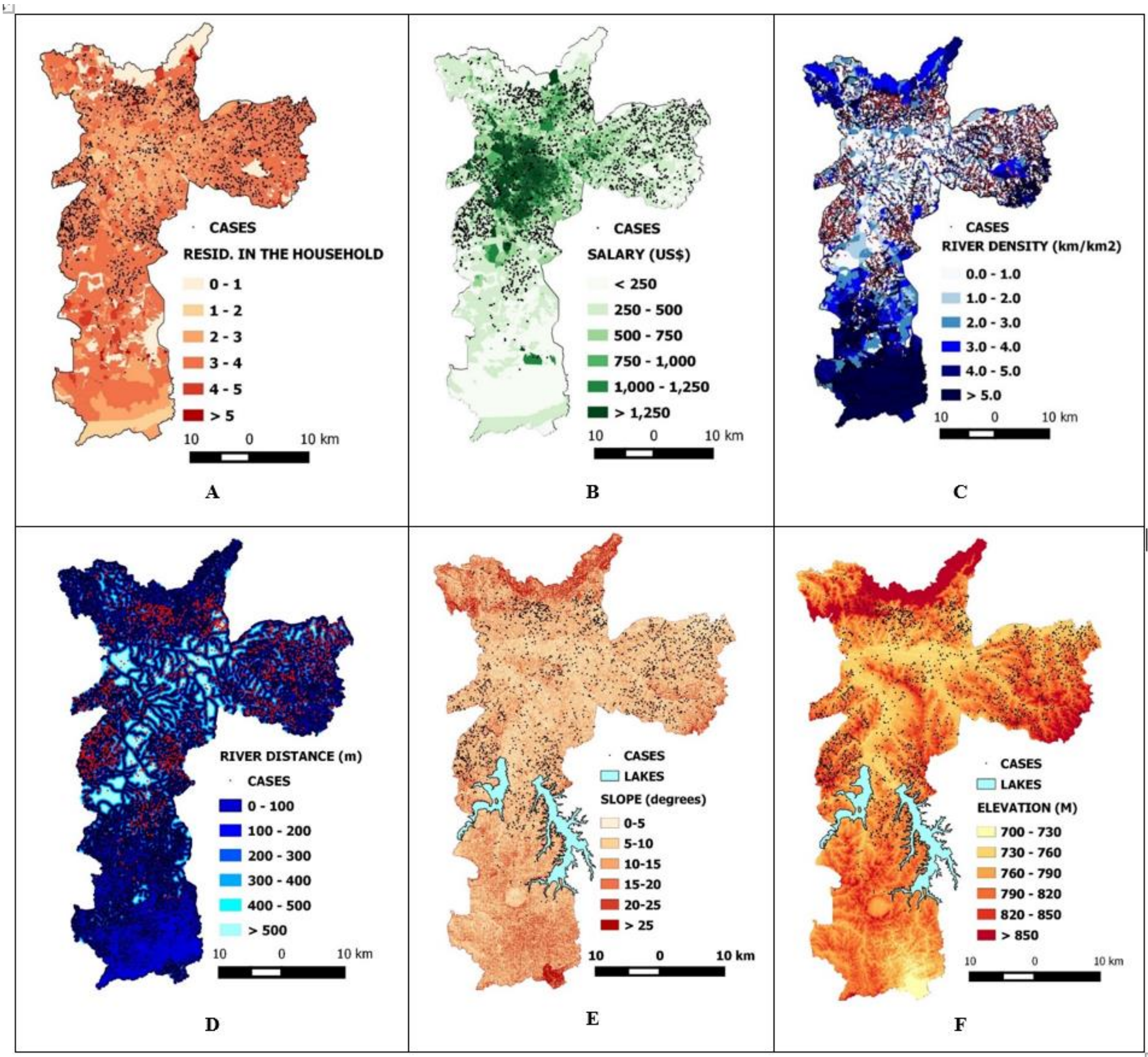

Figure 5. Spatial distribution of the leptospirosis cases that were documented from 1998 to 2007 in Sao Paulo in relation to the average number of residents in the household (A), average monthly salary (B), river density in the census sector (C), river distance (D), slope (E) and altitude (F).

4.3 - Statistical Analysis

The results of the statistical analysis are shown in Table 3 and Figure 6. In relation to the topographical variables, we found that the average slope in the case cells (SLO = 7.111°) was higher than that in the no-case cells (SLO=6.928°) ($p=0.1200$). The average altitude was higher in the case cells (ELV=775.07 m) than that in the no-case cells (ELV=774.86 m) ($p=0.8679$). In relation to the hydrographic variables, we observed that the distance to rivers in the case cells (RDI=141.80 m) was lower than that in the no-case cells (RDI=152.73 m) ($p=0.1025$). Furthermore, the river density in the census sector in the case cells (RDE=2.547 km/km^2) was higher than that in the no-case cells (RDE=2.309 km/km^2) ($p=0.0012$).

Table 3. Average values for the SAL, SLO, RDE, RDI, RES and ELV variables that were calculated for the case cell and no-case cell maps and their respective p-values that were estimated using the independent samples t-test.

	Case cells (n = 1,027)		No-case cells (n = 653)		
Variables	Average	95% CI	Average	95% CI	*p*
SAL (US$)	386.14	371.75 to 400.53	479.52	455.17 to 503.86	*<0.0001(1)*
RES	3.195	3.180 to 3.210	3.091	3.067 to 3.115	*<0.0001(1)*
SLO (°)	7.111	6.961 to 7.260	6.928	6.753 to 7.103	*0.1200*
RDE (km/km^2)	2.547	2.456 to 2.638	2.309	2.199 to 2.419	*0.0012(1)*
RDI (m)	141.80	133.63 to 149.97	152.73	142.44 to 163.022	*0.1025*
ELV (m)	775.07	773.53 to 776.60	774.86	772.86 to 776.85	*0.8679*

(1) *Differences between the average values that were statistically significant at p<0.005.*

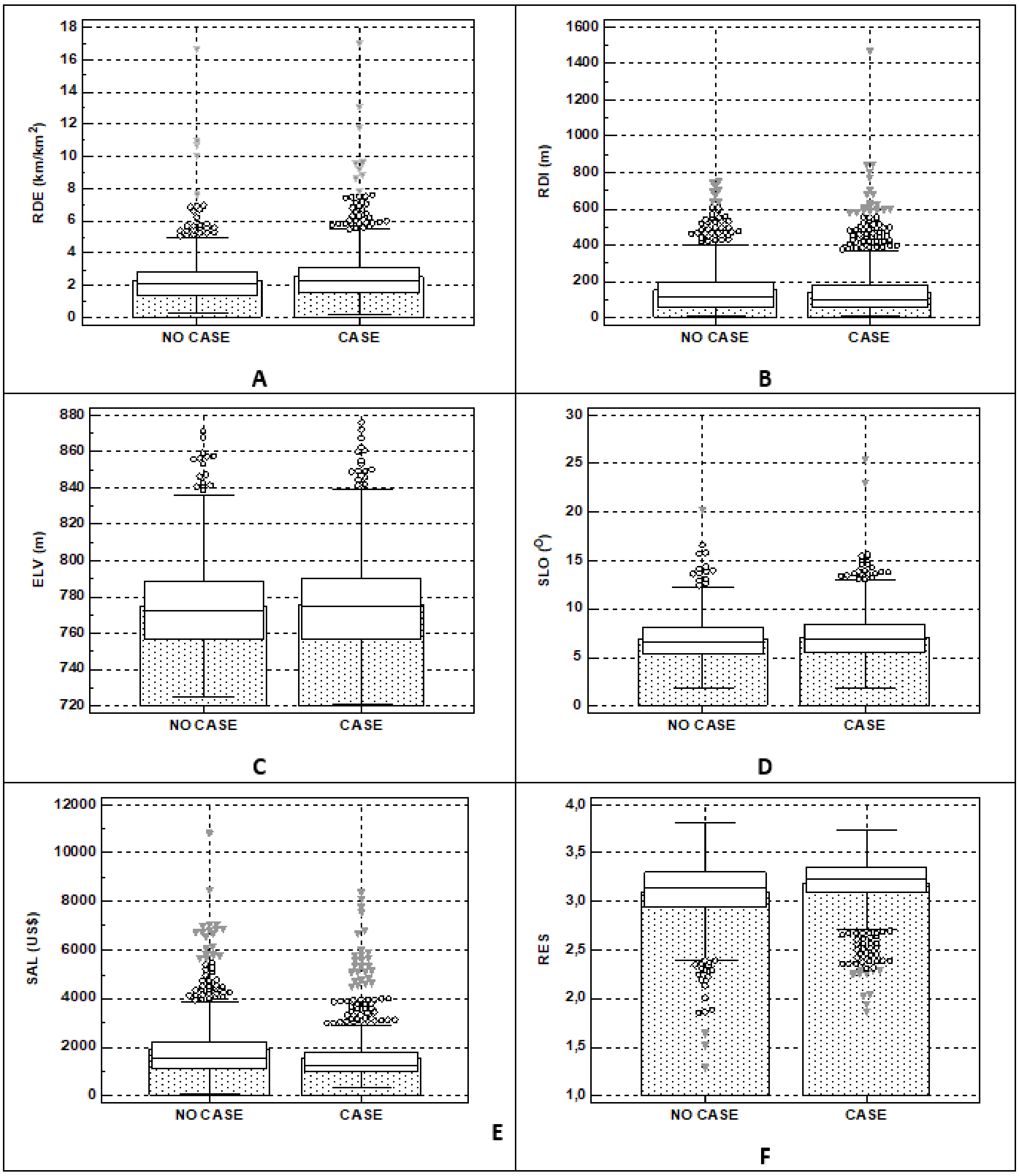

Figure 6. Data comparison graphs showing 95% confidence intervals for the means that were estimated for the case cells and no-case cells for the river density in the census sector (A), river distance (B), altitude (C), slope (D), average monthly salary (E) and average number of residents in the household (F).

The results from the socioeconomical variables showed that the average monthly salary that was calculated for the case cells (SAL=386.14 US$) was lower than that in the no-case cells (SAL=479.52 US$) ($p<0.0001$). With respect to the number of residents in the household, the average for the case cells (RES=3.195) was higher than it was for the no-case cells (RES=3.091) ($p<0.0001$).

5. DISCUSSION

Due to the great number of cases that were observed over the 10 years that were analysed, we assume that the areas corresponding to these five nuclei are leptospirosis epidemiological landscape units that are composed of microhabitats that have social and environmental characteristics favouring the contagion of the Leptospira bacteria.

Although we confirmed the positive spatial autocorrelation of leptospirosis case incidences in Sao Paulo (I=0.27) in our study, this index was not statistically significant. Soares et al (2010) found very similar values, i.e., I=0.26 for the incidence of cases during the dry period and I=0.19 during the wet period. This finding may indicate that during the rainy season, the disease probably spreads, decreasing the local agglomerates. However, during the dry period, the cases aggregate in the epidemiological landscape units.

The values that are found in our study for the average incidence rate of leptospirosis in the districts of Sao Paulo from the period from 1998-2007 varied from 0.25/100,000 (Vila Mariana district) to 4.31/100,000 (Se district). These values were relatively close to those that were obtained by Guimaraes et al. (2014) for Rio de Janeiro, Brazil (0.54 to 5.71/100,000) and by Sanchez-Montes et al (2015) for Mexico (0.08 to 11.75/100,000). However, our results were smaller than those that were found by Costa et al. (2014) for the slums of Salvador, Brazil (35.4/100,000).

The districts that are in the highest quartile of incidence rates are situated in the clusters that are mapped by Soares et al (2010) using Moran's index; these results are similar to the literacy rate, average monthly income and average number of residents by household. Although we confirmed the positive spatial autocorrelation of leptospirosis case incidences in Sao Paulo in our study (I=0.27), this index was not statistically significant. Soares et al. (2010) found very close values, i.e., I=0.26 for the incidence of cases during the dry period and I=0.19 during the wet period. This finding may indicate that during the rainy season, the disease probably spreads, decreasing the local agglomerates.

The leptospirosis cases that were used in our study were reported during both the dry and wet seasons. However, we know that the associations with the river density and river distance variables may be higher during the wet season than during the dry season. However, during the dry season, most cases occur in poor neighbourhoods with households that are distant from rivers (Soares et al., 2010).

Family salary was significantly lower in the case cells than in the no-case cells. The number of residents in the household was significantly higher in the case cells than in the no-case cells. The research that carried out by Soares et al. (2010) found a significant negative correlation between the incidence rate of leptospirosis and average income in Sao Paulo districts. Ayral et al. (2015) showed that rodent-transmitted disease risk is higher in the low-income neighbourhoods of Rhône, France.

The river distance was not significantly lower in the case cells than in the no-case cells. The river density in the census sector was significantly higher in the case cells than in the no-case cells. A higher frequency of cases occurred mainly in small hydrographic basins with high river densities in the census sector, which were occupied by slums and had populations that suffer because of the floods after rainstorms.

In relation to the river distance, our results were different from the findings of Sunaryo (2012), who studied the city of Semarang, Java. That author concluded that a higher frequency of cases was observed in households that were situated 50 to 300 metres from rivers. Similarly, in Sri Lanka, a higher risk for leptospirosis was associated with areas less than 400 metres from rivers (Robertson et al., 2012).

In the República, Santa Cecília, Bela Vista and Consolaçao districts, which are situated in the old central area of Sao Paulo city, we identified a census sector cluster with a high number of cases, but the sector is distant from rivers. In these districts, there are some blocks with multiple families residing in the same households with inadequate sanitary conditions (PMSP, 2019). Perhaps this fact contributed to the occurrence of a greater number of cases in districts where households are situated far from rivers. However, this suggestion needs to be investigated in the future.

6. CONCLUSIONS

We used geospatial techniques and GIS to study the relationships between the spatial distribution of leptospirosis and the socioeconomic and physical geographic variable maps in Sao Paulo, Brazil. This geographical approach permitted us to assess the relative importance of each spatial variable in the occurrence of leptospirosis. The vector techniques allowed us to analyse the punctual locations of the cases according to the polygonal geographical units, such as municipal districts and the census sectors. The use of vector techniques facilitated the estimation of the observed frequencies according to the socioeconomic class intervals, as well as the incidence rate mapping and calculation of the river density by census sector.

In addition, the use of the vector statistical grid of the Brazilian National Census of 2010, which was associated with the raster and vector data corresponding to the socioenvironmental variables, allowed us to obtain data on the population and statistical parameters that we used in the significance tests.

The average family salary, the number of residents in households and the river density in the census sector were more associated with leptospirosis than the distance from a river and the topographic variables (slope and altitude). The differences between the geospatial techniques that were used to map the socioeconomic (vector analysis) and physical geographic variables (raster analysis) may have influenced the lower statistical significance values for the topographical and hydrographical variables. In addition, the interpolation method that was used to generate surfaces may have underestimated the associations of these variables with the spatial distribution of leptospirosis cases.

ACKNOWLEDGEMENTS

We thank the National Council for Scientific and Technological Development of Brazil (CNPq) for funding this research, and the Metropolis Study Centre (CEM) and Health Surveillance Coordination of Sao Paulo (COVISA) for providing and publishing the epidemiologic data that were used in this research.

REFERENCES

Anselin, L. (2003). *GeoDa 0.9 user's guide*. Urbana, IL, Center for Spatially Integrated Social Science, http://www.csiss.org (Accessed 2016/01/10).

ASTER GDEM. (2014). *Global Digital Elevation Model version 2*. http://gdem.ersdac.jspacesystems.or.jp (Accessed 2014/10/06).

Ayral, F., Artois, J., Zilber, A.L., Widen, F., Pounder, K.C., Aubert, D., Bicout, D.J., and Artois, M. (2015). The relationship between socioeconomic indices and potentially zoonotic pathogens carried by wild Norway rats: a survey in Rhône, France (2010-2012); *Epidemiol. Infect*. 143, 586-599. DOI:10.1017/S0950268814001137.

Azevedo, R., and Correa, M. (1968). Considerações em torno da epidemia de leptospirose na cidade de Recife, em 1966. Aspectos epidemiologicos, laboratoriais e clínicos. *Revista do Instituto Adolfo Lutz*, 28: 85-111.

Barcellos, C., and Sabroza, E.P.C. (2001). The place behind the case: leptospirosis risks and associated environmental conditions in a flood related outbreak in Rio de Janeiro. *Caderno de Saúde Pública*, 17:59-67.

Caldas, E.M. (1979). Estudo epidemiologico de um surto de leptospirose na cidade de Salvador em maio de 1979. *Revista do Instituto Adolfo Lutz*, 39(1):35-94.

CEM. (2015). *Centro de estudos da metropole*. http://www.fflch.usp.br/centrodametropole (Accessed 2015/03/20).

Costa, F., Ribeiro, G.S., Felzemburgh, R.D.M., Santos, N., Fraga, D.B.M., and Araujo, W.N. (2014). Influence of household rat infestation on Leptospira transmission in the urban slum environment. *PLOS Neglected Tropical Diseases*, 8 (12) e3338.

Doukissas, L., Kalogirou, S. and Tsimbos, C. (2018). Spatial patterns of SMRS due to the virus A(H1N1)PDM09 during the pandemic in Greece in 2009 – 2010. *European Journal of Geography*, 9 (2):134-148.

Fenn, M.G.P., and Mac Donald, D. W. (1987). The contribution of field studies to stored product rodent control. In: LAWSON, T. J. (ed.). *Stored products pest control*. Thornton Heath (UK): British Council of Pest Control (BCPC) p. 107-113.

Ferreira, M. C. (2014). Geographical distribution of the association between El Niño South Oscillation and dengue fever in the Americas: a continental analysis using geographical information system-based techniques. *Geospatial Health*, *9*(1), 141-151. https://doi.org/10.4081/gh.2014.12.

Figueiredo, C.M., Mourao, A.C., Oliveira, M.A.A., Alves, W.R., Ooteman, M.C., Chamone, C.B., and Koury, M.C. (2001). Leptospirose humana no município de Belo Horizonte, Minas Gerais, Brasil: uma abordagem geografica. *Revista Brasileira de Medicina Tropical*, 34 (4):331-338.

GEOSAMPA, Prefeitura Municipal de Sao Paulo (2017). *Mapa digital da cidade de Sao Paulo*; www.geosampa.prefeitura.sp.gov.br (Accessed 2017/10/25).

Gracie, R., Barcellos, C., Magalhaes, M., Souza-Santos, R., and Barrocas, P.R.G. (2014). Geographical scale effects on the analyst of leptospirosis determinants; *Intern. J. Environ. Res. Public Health*, 11, 10366-20383. DOI10.3390/ijerph111010366.

Gubler, D.J. (1996). The global resurgence of arboviral diseases. *Trans. R. Soc. Trop. Med. Hyg*. 90:449–51.

Guimaraes, R.M., Cruz, O.G., Parreira, V.G., Mazoto, M.L., Vieira, J.D., and Asmus, C.I.R.F. (2014). Analise temporal da relaçao entre leptospirose e ocorrência de inundações por chuvas no município do Rio de Janeiro, Brasil, 2007-2012 (Temporal analysis of the relationship between leptospirosis and the occurrence of flooding due to rainfall in the city of Rio de Janeiro, Brazil, 2007-2012). *Ciência & Saúde Coletiva*, 19 (9):3683-3692.

Hartskeerl, R.A., Collares-Pereira, M., and Ellis, W.A. (2011). Emergence, control and re-emerging leptospirosis: dynamics of infection in the changing world. *Clin. Microbiol Infect*. 17:494-501.

IBGE, Instituto Brasileiro de Geografia e Estatística (2010). *Censo Nacional do Brasil*; www.ibge.gov.br/home (Accessed 2014/10/06).

Kraemer, M.U.G., Faria, N.R., and Reiner, R.C. et al. (2017). Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015-16: modelling study. *The Lancet*, 17: 330-338.

Levett, P.N. (2001). Leptospirosis. *Clinical Microbiology Reviews*, 14:296-326.

Luenam, A., Puttanapong, N. (201)9. Spatial and statistical analysis of leptospirosis in Thailand from 2013 to 2015. *Geospatial Health*, *14*(1). https://doi.org/10.4081/gh.2019.739.

Mahmood, S., Irshad, A., Nasir, J.M., and Sharif, F. (2019). Spatiotemporal analysis of dengue outbreaks in Samanabad town, Lahore metropolitan area, using geospatial techniques. *Environ Monit Assess*, 191: 55. https://doi.org/10.1007/s10661-018-7162-9

Masi, E., Pino, F.A., Santos, M.G.S., Genehr, L., Albuquerque, J.O.M., Bancher, A.M., Alves, J.C.M., and Santana, C.A. (2009). Fatores determinantes da infestaçao predial por roedores nas subprefeituras de Sao Paulo (Risk factors of urban rodent infestation in Sao Paulo boroughs). *Hygeia*, 5 (8):29-45.

MedCalc Statistical Software version 18.6. 2018. *MedCalc Software*, Ostend, Belgium; http://www.medcalc.org.

Pavloskiy, E.N. (1966.) *Natural nidality of transmissible diseases with special reference to the landscape ecology of zooanthroponses*. Urbana: Univ. Ill. Press

PMSP. (2007). Identificaçao e delimitaçao de areas prioritarias para controle da leptospirose em Sao Paulo (*Identification and Delimitation of Priority Areas for Leptospirosis Control in Sao Paulo Report*. Prefeitura Municipal de Sao Paulo. http://www.prefeitura.sp.gov.br (Accessed 2017/10/18).

PMSP. (2019). *Prefeitura Municipal de Sao Paulo*; Dados Abertos; Layer Cortiços; http://dados.prefeitura.sp.gov.br/dataset/corticos/resource/90f20793-527b-4291-a377-4041ee3e4e9b (Accessed 2019/04/03).

QGIS 2018 - *Quantum GIS Desktop* v. 3.02. www.gnu.org/licenses.

Raghavan, R.K., Brenner, K.M., Harrington, J.A., Higgins, J.J., and Harkin, K.R. (2013). Spatial scale effects in environmental risk-factor modelling for diseases. *Geospatial Health*, 7(2):169-182.

Raghavan, R.K., Brenner, K.M., Harrington, J.A., Higgins, J.J., Harkin, K.R. (2012). Evaluations of hydrologic factors for canine leptospirosis: 94 cases (2002-2009). *Preventive Veterinary Medicine*, 107:105-109.

Reis, R.B., Ribeiro, G.S., Felzemburg, R.D.M., Santana, F.S., Mohr, S., and Melendez, A. (2008). Impact of environment and social gradient on Leptospira infection in urban slums; *PLOS Neglected Tropical Diseases*; 2. DOI:10.1371/jornal.pntd.0000228.

Reisen, W.K. (2010). Landscape epidemiology of vector-borne diseases. *Annual Review of Entomology*, 55:461-483.

Robertson, C., Nelson, T.A., and Stephen, C. (2012). Spatial epidemiology of suspected clinical leptospirosis in Sri Lanka. *Epidemiology Infectious*, 140:731-743.

Sanchez-Montes, S., Espinosa-Martínez, D.V., Ríos-Muñoz, C.A., Berzunza-Cruz, M., and Becker, I. (2015). Leptospirosis in Mexico: epidemiology and potential distribution of human cases; *PLoS ONE*; 10 (7): e0133720. DOI: 10.1371/journal.pone.0133720.

SEADE, Secretaria de Planejamento e Desenvolvimento Regional, Governo do Estado de Sao Paulo. (2017). *Sistema estadual de analise de dados*, http://www.seade.gov.br/. Accessed 2017/10/10).

SIH, Ministerio da Saúde do Brasil. (2007). *Sistema de informaçao hospitalar*, Centro de Estudos da Metrpole (Accessed 2014/06/10).

Silva, A.R.M., Quadra, A.F., and Cordeiro, H.A. (1975). Aspectos epidemiologicos das leptospiroses humanas no Grande Rio, Brasil. *Boletin de La Oficina Sanitaria Panamericana*. 8:122-133.

Smith, M.J., Goodchild, M.F., and Longley, P.A. (2009). *Geospatial analysis: a comprehensive guide to principles, techniques and software tools*. Leicester, UK, Splint.

Soares, T.S.M., Latorre, M.R.D.O., Laporta, G.Z., and Buzzai, M.R. (2010). Analise espacial da leptospirose no município de Sao Paulo, SP, 1998 a 2006 (Spatial and seasonal analysis on leptospirosis in the municipality of Sao Paulo, Southeastern Brazil, 1998 to 2006). *Revista de Saúde Pública*, 44 (2):283-291.

Sunaryo, D.W. (2012). Mapping of leptospirosis risk factor based on remote sensing image in Tembalang, Semarang City, Central Java. *Health Science Indonesian*, Vol. 3, N. 1, 45-50.

Tachikawa, T., Kaku, M., Iwasaki, A., Gesch, D., and Oimoen, M. (2011). *ASTER Global Digital Elevation Model version 2 – summary of validation results*. NASA Land Processes Distributed Active. Archive Center and the Joint Japan-US ASTER Science Team, Tokyo.

Traxler, R.M., Callinan, L.S., Holman, R.C., Steiner, C., and Guerra, M. (2014). Leptospirosis-associated hospitalizations, United States, 1998-2009. *Emerging Infectious Diseases*, vol. 20, N. 8, 1273-1279.

Veja-Corridor, M.C., and Opadeyi, J. (2014). Hydrology and public health: linking human leptospirosis and hydrological dynamics in Trinidad, West Indies. *Earth Perspectives*, 1:3, 1-14.

Ward, M.P., Guptill, L.F., and Wu, C.C. (2004). Evaluation of environmental risk factors for leptospirosis in dogs: 36 cases (1997-2002). *Journal of American Veterinary Medical Association*, 225, 72-77.