DOI: 10.36558/rsc.v12i3.7921

Identification of risk areas using spatial clustering to improve dengue monitoring in urban environments

Gesiel Rios Lopes
Institute of Mathematical and Computer Sciences
University of São Paulo
São Carlos - SP - BR
gesielrios@usp.br

Roberto Fray da Silva
Institute of Advanced Studies
University of São Paulo
São Paulo - SP - BR
roberto.fray.silva@gmail.com

Karina Jorge Pelarigo
Department of Nursing
Federal University of São Carlos
São Carlos - SP - BR
karousfcar@hotmail.com

Mellina Yamamura
Department of Nursing
Federal University of São Carlos
São Carlos - SP - BR
mellina@ufscar.br

Alexandre C. B. Delbem
Institute of Mathematical and Computer Sciences
University of São Paulo
São Carlos - SP - BR
acbd@icmc.usp.br

Antonio Mauro Saraiva
Polytechnic School
University of São Paulo
São Paulo - SP - BR
saraiva@usp.br

ABSTRACT

Monitoring the occurrence and spread of epidemics is essential for improving decision-making and developing better public policies in urban environments. Besides temporal aspects, it is also essential to evaluate risk areas. However, only a few works in the literature apply spatial analysis of dengue epidemics in Brazil due mainly to a lack of data availability. Additionally, few methodologies available allow for identifying risk areas considering spatial aspects. The main objective of this work was to identify spatial clusters of risk for dengue cases according to the social vulnerability of each area. This constitutes a powerful tool for effective epidemiological and urban management. This work carries out an ecological study that considered dengue cases in São Carlos-SP, Brazil, in the years 2018, 2019 and 2020. The spatial scan technique was applied to classify the risk areas, considering the relative risk (RR) with a confidence interval of 95% (CI95%:) and the São Paulo Social Vulnerability Index (IPVS) to characterize these areas. Three clusters were identified in 2018, with high risk relative (RR=28.86), twenty clusters were identified in 2019, with high risk relative (RR=36.26) and five clusters were identified in 2020, with high risk relative (RR=23.32). The highest risk was located in a region with high vulnerability, and the second was in a region with very low vulnerability. These results provide information that allows the targeting of specific control actions from the early detection of cases in places with greater dengue transmissibility.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CCS Concepts

•Applied computing → Health care information systems; **•Information systems → Geographic information systems;**

Keywords

Spatial clustering; Dengue; Vulnerability; Risk areas classification; Decision-making

1. INTRODUCTION

Dengue is an acute viral infection mainly transmitted by the *Aedes aegypti* mosquito. It is the most frequent arboviral disease worldwide, occurring mainly in the tropics and subtropics. It has an estimated burden of 390 million cases annually, of which 96 million cases manifest symptomatically, 2 million cases develop severe disease, and 21,000 lead to deaths [8].

The worldwide spread of dengue is a complex problem. This can be accelerated by several factors, such as climate change, population growth, rapid and unplanned urbanization, movement of people for commerce, tourism, natural disasters, and weaknesses in public health and vector control programs. From 2010 to 2019, more than 16 million dengue cases were reported across the American continent, and about 10 million cases (62%) were reported in Brazil alone [5, 8].

Dengue has a wide geographic distribution in this country. Despite the intensification of control measures, there has been an increase in severe cases, hospitalizations, and deaths in recent years. One of the causes identified as responsible is directly related to the rampant growth of cities, accompanied by the lack of awareness of the population in the elimination of mosquito breeding sites, which can be any container that accumulates rainwater [7, 5].

This work aims to identify risk clusters through spatial scanning statistics, a methodology developed by [11]. The notifications of dengue cases were considered as inputs. The main objective is to provide critical information for decision-making via the visualization of the spatial distribution of dengue cases.

This would allow health authorities to identify priority areas quickly and better direct efforts to control dengue in urban environments [9, 7, 5]. Similar processes have been proposed to combat malaria in the Brazilian Amazon [6] and tuberculosis in the municipality of São Carlos-SP [3]. It is vital to observe that the methodology used could be replicated for any urban environment and that additional available inputs could also be considered.

The rest of the paper is organized as follows. In Section 2 contains the methodology used. Section 3 contains the main results and discussions, including a description of the case study and the inputs used. Finally, Section 4 concludes the work, including recommendations for future works.

2. MATERIAL AND METHODS

The methodology used in this work follows the methodologies used by [6], [3], and [11], applied for the study of dengue in urban environments. It can be divided into five main steps:

1- Data gathering from official sources for 2019, considering both the reported and confirmed cases of dengue, the presence of breeding sites with larvae, and the vulnerability index for the municipality of São Carlos-SP. Section 3.1 describes the study area, and section 3.2 further describes the data collected;

2- Data fusion, considering the census sectors as the spatial unit. This was essential to aggregate the data of breeding sites and reported and confirmed cases of dengue, as they represented points, and the vulnerability index was calculated by area for each census sector. Section 3.2 describes in depth the data fusion methodology used;

3- Exploratory spatial data analysis, including implementing the Global and Local Moran indicators for evaluating the spatial correlation of each input and the spatio-temporal scanning technique output. Section 3.2 further describes these indicators and their implementation;

4- Implement the spatio-temporal scanning technique, developed by [11]. This method is essential to obtain insights from spatio-temporal data. It is widely used in the literature to cluster spatial data and generate visualizations that can improve decision-making. Section 3.3 further describes the implementation of this technique;

5- Calculate the relative risk (RR) and evaluate the spatial clusters generated. Section 3.3 contains the results of this analysis.

3. RESULTS AND DISCUSSIONS

This section contains the main results of this work, as well as discussions related to their importance from the point of view of the decision-makers. It is divided into the following subsections: 3.1 describes the study area; 3.2 contains the data used and the main results of the exploratory data analysis conducted; and 3.3 presents the main results of the spatial clustering implemented.

3.1 Study area

As observed in section 1, the main objective of this work is to conduct a geographic and ecological study of dengue in São Carlos-SP, Brazil. The unit of analysis used was the census tracts of the urban area of the municipality. This is a medium-sized city in the interior of São Paulo state. Located in the central-eastern region of the state (Figure 1), at the coordinates 22°1'4" South latitude and 47°53'27" West latitude, São Carlos had a total territorial area of 1,136,907 km^2, an average altitude of 856 meters, a population density of 195.15 inhabitants/km², and a total population of 221,950 inhabitants in 2010. Regarding socioeconomic aspects, the municipality had a Gini index of 0.63, an Human Development Index (HDI) of 0.805, and a gross domestic product of R$6,712,498.00 for the same year [10]. Therefore, it can be considered a wealthy city by Brazilian standards.

Figure 1: Municipality of São Carlos-SP and its urban perimeter.

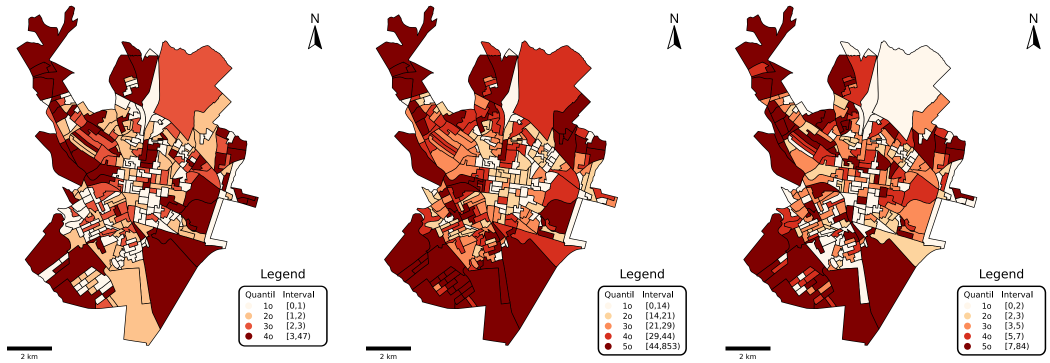
3.2 Data source

The data collected for this work were: (i) all reported and confirmed cases of dengue from the Information System on Notifiable Diseases-Sinan Dengue/Chikungunya of residents of the urban area of the municipality of São Carlos-SP from January 1 to December 31, of the years 2018, 2019 and 2020; and (ii) the official social vulnerability index, calculated for each census sector (`https://ipvs.seade.gov.br/view/index.php`).

For the data analysis step, the data was initially georeferenced for the notifications of dengue cases and the places visited by endemic agents. For this procedure, the authors developed a python *script* using the Google Maps geolocation API to obtain the respective geographic coordinates (latitude and longitude) of the notified addresses in the cases and the visits of endemic agents. After geolocation, the cases and actions of endemic agents were aggregated with the digital mesh of the IBGE's urban census sectors of São Carlos through a *join spatial* function, which crosses the layers. Cases with unspecified addresses, duplicates, and residents in the rural area of the municipality were excluded from the analysis.

The usual way of presenting aggregated data by area is through choropleth maps with the spatial pattern of the phenomenon [4]. Thus, Figure 2(a), Figure 2(b) and Figure 2(c) bring the choropleth representations of the spatial distribution of dengue cases in 2018, 2019 and 2020, respectively. For better data visualization, aggregated data from all maps were sliced into groups according to their quintiles to form classes with an approximate number of assigned features.

To ensure that the data aggregations in the census sectors represent a phenomenon from a spatial point of view,



(a) (b) (c)

Figure 2: Choropleth representation of the spatial distribution of dengue cases in: (a) 2018, (b) 2019 and (c) 2020.

the spatial autocorrelation was calculated using the Global and Local Moran indicators [16, 12] considering only the first neighborhood level. The Global Moran indicator was used as a test whose null hypothesis is data independence. Global indicators such as Moran I provide a single value to measure spatial association for the entire dataset, which helps characterize the region. The global Moran index is given by Equation 1.

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(z_i - \bar{z})(z_i - \bar{z})}{\sum_{i=1}^{n}(z_i - \bar{z})^2} \quad (1)$$

where n is the number of areas, z_i is the value of the attribute considered in the area i, z is the average of the assigned values in the study region, and w_{ij} the elements from the normalized spatial proximity matrix [16, 12]

Moran's I for dengue cases in 2018, 2019 and 2020 was 0.0621, 0.1064 and 0.1086, respectively. Given the *z-score* of 2.178 for dengue cases in 2018, 4.655 in 2019 and 3.863 in 2020 and for both analyzes, there is a probability less than 1% that the pattern clustering may be an unexpected result. Thus, the dataset's spatial distribution of high and low values is more spatially clustered than expected if the underlying spatial processes were random, thus rejecting the null hypothesis [4]. This is important because it directly points out the existence of spatial correlation. This aspect must be considered in designing and implementing public policies to monitor and control dengue in urban environments.

Once the global dependence was verified, the Local Spatial Association Index (LISA) was calculated. LISA is a decomposition of Moran's I, in which it is possible to analyze the local pattern of spatial data. LISA can be expressed for each area i from the normalized values of the attribute's z_i by Equation 2.

$$I_i = \frac{z_i \sum_{j=1}^{n} w_{ij} z_j}{\sum_{j=1}^{n} w_{ij} z_j^2} \quad (2)$$

Based on the LISA, the census tracts with the aggregated data are positioned in the quadrants of the Moran scatterplot as follows: (i) Q1 (high-high), census tracts where the attribute value and the mean value of neighbors are above average as a whole, and which are considered to be the highest priority for intervention; (ii) Q2 (low-low), the attribute value and the mean of the neighbors are below the mean of the set; (iii) Q3 (high-low), the attribute value is greater than that of the neighbors and the mean of the neighbors is less than the set; and (iv) Q4 (low-high), the attribute value is lower than that of the neighbors and the average of the neighbors is higher than the average of the set.

Based on the LISA analysis, municipalities classified as high-low and low-high have intermediate priority [15, 2]. Therefore, this analysis is vital to present both the most critical sectors, which must be considered urgent, and the least critical sectors, which could be

Figure 3(a), Figure 3(b) and Figure 3(c) presents the result of the analysis of the clusters for dengue cases in 2018, 2019 and 2020, respectively. It is possible to observe that in 2018 the census sectors classified as Q1 (high-high), therefore the most critical, are located in the northwest region of the municipality. In 2019 and 2020, the census sectors classified as Q1 are located in the northwest and southern regions of the municipality.

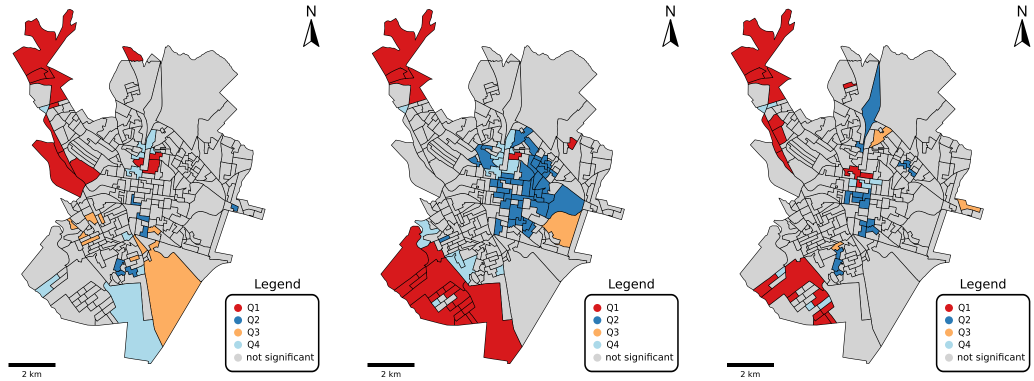
Two important observations related to those results are: (i) decision-makers could use this information to target control initiatives in the northwest and southern regions of the city; and (ii) this result complements the one observed by analysis of the Global Moran values. Therefore, both analyses should provide decision-makers with potentially relevant information for public policy design and monitoring and control activities.

3.3 The spatio-temporal scanning technique

The spatial scanning statistical technique was developed by [11], as already mentioned, in order to identify and locate risk clusters present in a given study region. For this analysis, separate spreadsheets were created to analyze dengue cases, considering the code of the census sector and the number of occurrences of each event.

The risk clusters are graphically identified by circular windows with a variable radius around the centroids of each census sector, for which the expected number of occurrences within the circle is calculated. The region delimited by the analysis window, called region z, can constitute a cluster if the value found is higher or lower than expected. This procedure is performed in all centroids under analysis [14].

Thus, the null hypothesis (H_0) against the alternative hypothesis (H_1) was tested differently among the diseases, highlighting that H_0 assumes that there are no clusters areas for dengue cases, that is, the population has the same



(a) (b) (c)

Figure 3: Cluster analysis according to the Moran'I Local of dengue cases in: (a) 2018, (b) 2019 and (c) 2020.

probability of contracting dengue case and H_1 assumes that one or more regions z are areas in which there would be the greater or lesser probability of contracting the diseases, compared to the that are outside that area.

In order to identify purely spatial clusters in which the distribution is heterogeneous. The events are rare in relation to the population. The discrete Poisson model was used with requirements of non-overlapping geographical clusters, clusters with a circular shape, 999 replications, and the size of the exposed population stipulated by the Gini coefficient released by the software itself. In this model, the number of cases was compared to the baseline population data, and each unit's expected number of cases was proportional to the population at risk [1].

The relative risk (RR) of each cluster will be calculated, allowing the comparison of information in different areas, indicating the intensity of occurrence of dengue cases in the municipality of São Carlos. It is noteworthy that the RR will be defined as the risk of having dengue in a risk area of the municipality in relation to the risk of having dengue outside this area.

Areas with *p-value* < 0.05 were considered statistically significant. The confidence interval was calculated and estimated at 95% [3]. Identifying the clusters' RR provides the comparison of information in distinct areas, as the effects of different populations are disregarded, thus resulting in the intensity of occurrence of the phenomenon under analysis throughout the study area. The values resulting from this calculation will be named high risk when the cluster RR is greater than one ($RR > 1$), and low risk when less than one ($RR < 1$) [3].

The cluster detection analyzes were performed using *software* SaTScan, a free *software* widely used by important Centers for Health Studies [1], version 10.1.

To describe the social vulnerability of the clusters found, data from the Fundação Sistema Estadual de Análise de Dados (SEADE), referring to the São Paulo Social Vulnerability Index (IPVS) for 2010, were used. This index classifies the census sectors based on a combination of the demographic and socioeconomic dimensions and identifies the specific factors that produce the deterioration of living conditions in a community, helping to define priorities for the care of the most vulnerable population[18].

The IPVS incorporates the following indicators: number of inhabitants; average nominal income of households; the average age of heads of households; percentage of heads of households under 30 years of age, female heads of households under 30 years of age, and the share of children under six years of age, over the denominator of the total inhabitants of each of these segments [18], characterizing the census sectors in seven groups: Group 1–extremely low vulnerability; Group 2–very low vulnerability; Group 3–low vulnerability; Group 4–medium vulnerability; Group 5–high vulnerability; Group 6–very high vulnerability and Group 7–very high vulnerability. Figure 4 shows the choropleth representation of the IPVS distribution for the city of São Carlos-SP

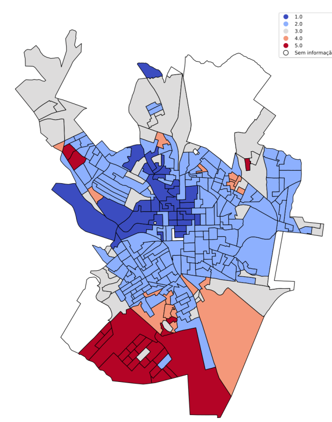


Figure 4: choropleth representation of the IPVS distribution for the city of São Carlos-SP.

3.4 Results

Sinan Dengue/Chikungunya reported 762, 10451 and 1650 dengue cases in 2018, 2019 and 2020, respectively, in the municipality of São Carlos.

Applying the spatial scanning statistics for dengue cases, three statistically significant clusters were detected in 2018, twenty statistically significant cluster in 2019 and five statistically significant cluster in 2020, as can be seen in Figure 5(a), Figure 5(b) and Figure 5(c), respectively. Table 1, Table 2 and Table 3 show the characteristics of the three statistically significant clusters with the highest risk for dengue, according to the temporal scan, in the municipality of São Carlos-SP in the years 2018, 2019 and 2020, respectively.

Table 1: Characteristics of the three clusters statistically significant in terms of risk for dengue cases, according to the spatial scan, in the municipality of São Carlos-SP in 2018.

	Cluster 1	Cluster 2	Cluster 3
Number of census sectors	1	1	52
Population	553	527	42612
Number of cases	45	17	194
Number of expected cases	1.66	1.74	142.01
RR	28.86	10.01	1.51

Table 2: Characteristics of the three clusters statistically significant in terms of risk for dengue cases, according to the spatial scan, in the municipality of São Carlos-SP in 2019.

	Cluster 1	Cluster 2	Cluster 3
Number of census sectors	1	1	1
Population	553	138	80
Number of cases	831	192	33
Number of expected cases	24.97	6.07	3.72
RR	36.26	32.26	8.90

Table 3: Characteristics of the three clusters statistically significant in terms of risk for dengue cases, according to the spatial scan, in the municipality of São Carlos-SP in 2020.

	Cluster 1	Cluster 2	Cluster 3
Number of census sectors	1	1	1
Population	597	138	911
Number of cases	84	20	19
Number of expected cases	3.82	0.92	6.01
RR	23.32	22.15	3.19

The municipality of São Carlos, as well as other industrial centers in Brazil, suffered significant urbanization in its peripheral regions [13]. A significant part of this growth occurred in the form of subdivisions of a precarious standard and concentrated in the low-income population, in the southeastern and southern sectors of the city [17].

When looking at Figure 5(a), Figure 5(b) and Figure 5(c), it is possible to detect that the census sectors that composed the purely spatial risk clusters come from the Cidade Aracy neighborhood, also located in the southern region of the city, and from the Jardim Tangará neighborhood, this one in the northeast region of São Carlos, the region with high social vulnerability, according to the IPVS.

On the other hand, the present analysis resulted in a cluster of low relative risk for dengue cases, with these census sectors located in the Vila Prado, Vila Boa Vista, Vila Carmem, and Jardim Beatriz neighborhoods, located in the southwest region of the municipality. These locations had an incidence rate below average, i.e., the number of space cases was lower than in any other region in the municipality, constituting areas of protection for infection by dengue cases.

It is worth mentioning that the municipality of São Carlos is considered a high-tech hub, a reference for the state of São Paulo, with an IDH, Gini index and poverty incidence better than the state average, which may explain the absence of classified census sectors. as very high vulnerability.

As a limitation of this research, it is highlighted that in ecological studies, the identified results cannot be interpreted at the individual level. Undoubtedly, the spatial scan statistics contributed to exposing the dengue scenario in São Carlos and the presence of geographic areas of the municipality that are more susceptible to illness and need specific actions to control the disease.

4. CONCLUSIONS AND FUTURE WORKS

With the use of the spatial scanning technique, it was possible to more accurately identify high-risk areas for contracting the dengue virus. It was a tool to help epidemiological surveillance priority areas for intervention to combat dengue.

Furthermore, looking at the clusters in more detail can lead to developing community education and awareness programs to improve dengue prevention. However, further studies are needed to better validate spatial *clusters* as a dengue surveillance method.

As future work, it is intended to carry out new studies with analytical designs capable of verifying, with more excellent reliability, possible associations between the areas of most significant risk for illness by dengue and Spatio-temporal analyses.

5. ACKNOWLEDGMENTS

The authors of this work would like to thank the Center for Artificial Intelligence (C4A-IUSP) and the support from the São Paulo Research Foundation (FAPESP grant #2019/07665-4, #2020/16578-5) and from the IBM Corporation. We would like also to thank the CEPID-CeMEAI-ICMC-USP (CEPID, FAPESP grant #2013/07375-0); the support of José Galizia Tundisi from the Science, Technology and Economic Development Department from São Carlos Municipality.

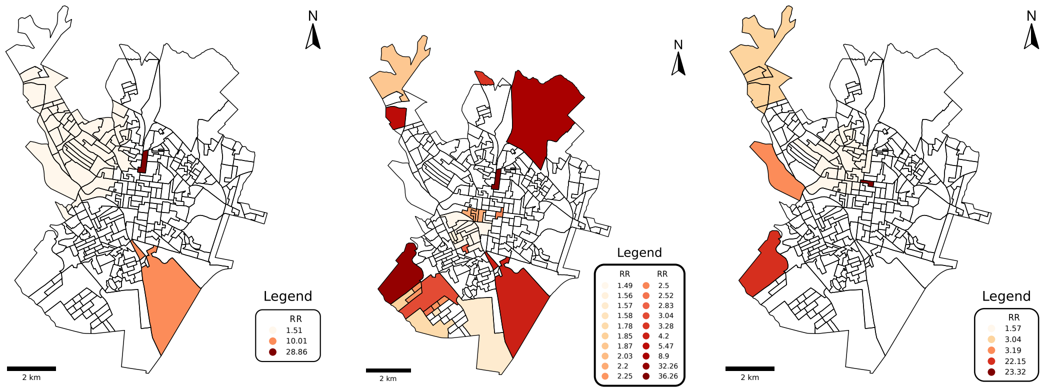
6. ADDITIONAL AUTHORS

Additional authors: Denise Scatolini (Center for Disease Control and Prevention, São Carlos Municipality, email: `denise.scatolini@saocarlos.sp.gov.br`) and Filippo Ghiglieno (Physics Department, Federal University of São Carlos, email: `filippo.ghiglieno@df.ufscar.br`).

7. REFERENCES

[1] L. S. Alves, D. T. Dos Santos, M. A. M. Arcoverde, T. Z. Berra, L. H. Arroyo, A. C. V. Ramos, I. S. De Assis, A. A. R. De Queiroz, J. B. Alonso, J. D. Alves, et al. Detection of risk clusters for deaths due to tuberculosis specifically in areas of southern brazil where the disease was supposedly a non-problem. *BMC infectious diseases*, 19(1):1–13, 2019.

[2] A. L. Andrade, A. M. V. Monteiro, C. Barcellos, E. Lisboa, L. M. W. Acosta, M. C. d. M. Almeida, M. R. V. Brito, M. S. Carvalho, M. A. d. Santos,



(a) (b) (c)

Figure 5: Cluster analysis according to the Moran'I Local of dengue cases in: (a) 2018, (b) 2019 and (c) 2020.

O. Cruz, et al. Introdução à estatística espacial para a saúde pública. 2007.

[3] L. H. Arroyo, M. Yamamura, S. T. Protti-Zanatta, A. P. B. Fusco, P. F. Palha, A. C. V. Ramos, S. A. Uchoa, and R. A. Arcêncio. Identificação de áreas de risco para a transmissão da tuberculose no município de são carlos, são paulo, 2008 a 2013. *Epidemiologia e Serviços de Saúde*, 26:525–534, 2017.

[4] G. Câmara, M. S. Carvalho, O. G. Cruz, and V. Correa. Análise espacial de áreas. *Análise espacial de dados geográficos*, 2, 2004.

[5] M. R. Canal, E. R. d. S. Ferreira, C. F. Estofolete, A. M. Dias, C. Tukasan, A. C. Bertoque, V. D. Muniz, M. L. Nogueira, and N. S. d. Silva. Spatiotemporal-based clusters as a method for dengue surveillance. *Revista Panamericana de Salud Pública*, 41:e162, 2018.

[6] N. S. Da Silva, M. da Silva-Nunes, R. S. Malafronte, M. J. Menezes, R. R. D'Arcadia, N. T. Komatsu, K. K. Scopel, É. M. Braga, C. E. Cavasini, J. A. Cordeiro, et al. Epidemiology and control of frontier malaria in brazil: lessons from community-based studies in rural amazonia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 104(5):343–350, 2010.

[7] D. P. O. de Melo, L. R. Scherrer, and Á. E. Eiras. Dengue fever occurrence and vector detection by larval survey, ovitrap and mosquitrap: a space-time clusters analysis. *PloS one*, 7(7):e42125, 2012.

[8] R. F. do Carmo, J. V. J. S. Júnior, A. F. Pastor, and C. D. F. de Souza. Spatiotemporal dynamics, risk areas and social determinants of dengue in northeastern brazil, 2014–2017: an ecological study. *Infectious Diseases of Poverty*, 9(1):1–16, 2020.

[9] E. L. Estallo, A. E. Carbajo, M. G. Grech, M. Frías-Céspedes, L. López, M. Lanfri, F. F. Ludueña-Almeida, and W. R. Almiron. Spatio-temporal dynamics of dengue 2009 outbreak in córdoba city, argentina. *Acta tropica*, 136:129–136, 2014.

[10] IBGE. Instituto brasileiro de geografia e estatística cidades - são calos. https://cidades.ibge.gov.br/brasil/sp/sao-carlos/panorama, 2022.

[11] M. Kulldorff and N. Nagarwalla. Spatial disease clusters: detection and inference. *Statistics in medicine*, 14(8):799–810, 1995.

[12] H. Li, C. A. Calder, and N. Cressie. Beyond moran's i: testing for spatial dependence based on the spatial autoregressive model. *Geographical analysis*, 39(4):357–375, 2007.

[13] R. P. Lima. *O processo e o (des) controle da expansão urbana de São Carlos (1857-1977)*. PhD thesis, Universidade de São Paulo, 2007.

[14] S. E. d. F. Lucena and R. M. d. Moraes. Detecção de agrupamentos espaço-temporais para identificação de áreas de risco de homicídios por arma branca em joão pessoa, pb. *Boletim de Ciências Geodésicas*, 18:605–623, 2012.

[15] A. M. V. Monteiro, G. Câmara, M. Carvalho, and S. Druck. Análise espacial de dados geográficos. *Brasília: Embrapa*, 2004.

[16] P. A. Moran. Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23, 1950.

[17] L. Schenk, M. Fantin, and R. Peres. A revisão do plano diretor da cidade de são carlos e as novas formas urbanas em curso. *Anais do X Colóquio Quapá-SEL: produção e apropriação dos espaços livres e da forma urbana*, 2015.

[18] SEADE. Fundação sistema estadual de análise de dados. Índice paulista de vulnerabilidade social. 2010.