

## ANÁLISE DE AGRUPAMENTO ESPACIAL DA LEPTOSPIROSE NO BRASIL

Alice Nardoni Marteli<sup>1</sup>, Laurindo Antonio Guasselli<sup>2</sup>, Décio Diament<sup>3</sup>, Gabriele Ozório Wink<sup>4</sup> e Vitor Vieira Vasconcelos<sup>5</sup>.

<sup>1</sup>Universidade Federal do Rio Grande do Sul (UFRGS), Programa de Pós-Graduação em Sensoriamento Remoto (PPGSR) – Porto Alegre (RS), [anm.nardoni@gmail.com](mailto:anm.nardoni@gmail.com); <sup>2</sup>UFRGS, PPGSR, [laurindo.guasselli@ufrgs.br](mailto:laurindo.guasselli@ufrgs.br); <sup>3</sup>Instituto de Infectologia Emilio Ribas (IIER) – São Paulo (SP), [ddiament@terra.com.br](mailto:ddiament@terra.com.br); <sup>4</sup>UFRGS, Instituto de Geociências, [gabrieleowink@gmail.com](mailto:gabrieleowink@gmail.com); <sup>5</sup>Universidade Federal do ABC (UFABC) – São Paulo (SP), [vitor.v.v@gmail.com](mailto:vitor.v.v@gmail.com).

### RESUMO

A leptospirose é um problema de saúde pública no Brasil. Este artigo analisou o agrupamento espacial da leptospirose no Brasil de 2007 a 2017. Técnicas de agrupamento, incluindo indicadores de associação espacial, foram utilizadas para análise e avaliação da distribuição espacial da doença. Os resultados mostram um *Global Moran's I* de 0,381 ( $p$ -value = 0.001) no período avaliado (2007-2017). O principal padrão observado mostra LISA clusters HH (High - High) no Norte e Sul do Brasil. O resultado da análise do agrupamento da leptospirose pode sugerir a órgãos gestores que recursos de saúde devem ser alocados para as áreas com aglomeração de leptospirose.

**Palavras-chave** — Sistema de Informação Geográfica, Detecção de aglomerados, Estudo Ecológico.

### ABSTRACT

*Leptospirosis is a public health problem in Brazil. This article analyzed the spatial clustering of leptospirosis in Brazil from 2007 to 2017. Clustering techniques, including spatial association indicators, were used to analyze and evaluate the spatial distribution of the disease. The results show a Global Moran's I of 0.381 ( $p$ -value = 0.001) in the evaluated period (2007-2017). The main pattern observed shows HH (High - High) LISA clusters map in the North and South of Brazil. The result of the leptospirosis clustering analysis can suggest to management agencies which health resources should be allocated to areas with agglomeration of leptospirosis.*

**Key words** — Geographic Information System, Cluster detection, Ecological study.

### 1. INTRODUÇÃO

A leptospirose é uma doença bacteriana transmitida por meio do contato com a urina de roedores. De acordo com a Organização Mundial da Saúde [1], o número de casos em humanos no mundo não é preciso, mas estima-se que a incidência em climas tropicais úmidos varia entre 10 e 100

por 100.000 pessoas, aumentando durante os surtos em grupos de risco de alta exposição. Além disso, a doença é subnotificada pela dificuldade em distinguir os sinais clínicos de outras doenças endêmicas [2].

Casos de leptospirose são registrados em várias partes do mundo. No Brasil está entre um dos problemas mais graves da saúde pública [3]. Dos 5.570 municípios, 2.600 tiveram casos confirmados de leptospirose entre 2007 e 2017, com uma média anual de 3.846 casos e incidência de 1,9 a cada 100 mil habitantes [4].

Estudos epidemiológicos têm explorado ferramentas de mapeamento da leptospirose em diferentes escalas [5–7] com o intuito de trazer novas abordagens ao processo saúde-doença da leptospirose. Entre as diversas técnicas de análise exploratória espacial utilizadas para detectar aglomerados, destaca-se o Índice de Moran, amplamente utilizado na área de saúde [2,8–10].

Assim, esse trabalho teve como objetivo analisar a presença de aglomerados e sua distribuição espacial da leptospirose no Brasil (2007-2017).

### 2. MATERIAL E MÉTODOS

Os dados de leptospirose foram obtidos do Ministério da Saúde pelo Sistema de Informação de Agravos de Notificação (SINAN) através do Departamento de Informática do Sistema Único de Saúde [11].

A taxa de incidência foi calculada utilizando dados de população obtidos a partir da estimativa da população residente por município e ano [12]. Para tal, foram considerados o número total de casos confirmados por município de residência e a população por município.

No GeoDa foram aplicadas técnicas de autocorrelação espacial para analisar e avaliar a distribuição espacial da leptospirose. Foi utilizado o método de estatística espacial *Global Moran's I* e *Local Indicators of Spatial Association* comumente aplicados na temática de leptospirose [2,10].

O *Global Moran's I* avalia se o atributo associado a localização espacial tem um padrão agrupado, disperso ou aleatório [13]. Para cada valor ele subtrai a média, criando um desvio da média. Os valores de desvio são multiplicados

para criar um produto cruzado [14], tal que para uma observação  $i$ :

$$I = \frac{\sum_i \sum_j w_{ij} z_i \cdot z_j / S_0}{\sum_i z_i^2 / n}$$

where  $w_{ij}$  são os elementos da matriz de pesos espaciais da prevalência de leptospirose nos municípios brasileiros  $i$  e  $j$ ;  $S_0 = \sum_i \sum_j w_{ij}$  é a soma de todos os pesos espaciais;  $n$  é o número total de casos confirmados; e  $z_i$  é o desvio de um atributo para o recurso  $i$  da sua média expresso como  $z_i = x_i - \bar{x}$ , onde  $\bar{x}$  é a média da variável  $x$  [15].

No GeoDa, o Global Moran's I é executado no menu *Space*, opção *Univariate Moran's I*. Como matriz de peso espacial foi utilizada a opção "*Queen contiguity*", de primeiro grau, recomendado para polígonos irregulares [16], como é o caso dos limites políticos municipais. Foi executada uma correção topológica para assegurar a contiguidade na construção da matriz de vizinhança, pois alguns municípios têm dois polígonos como limite territorial. Municípios insulares como Fernando de Noronha (PE) e Ilha Bela (SP) não foram relacionadas por *default* e, por isso foi necessário alterar a matriz de vizinhança.

A significância da estatística do teste foi obtida por meio da opção *Randomization*, com 999 permutações. A inferência de permutações calcula uma distribuição de referência para a estatística sob a hipótese nula de aleatoriedade espacial permutando aleatoriamente os valores observados (vizinhança) sobre o conjunto de dados gerados aleatoriamente [15]. No entanto, o agrupamento resultante da estatística *Global Moran's I* não fornece uma indicação da localização dos agrupamentos. Diante dessa limitação, Anselin [17] propôs o *Local Indicators of Spatial Association* (LISA) que fornece uma estatística para cada local com uma avaliação de significância e estabelece uma relação proporcional entre a soma das estatísticas locais e uma estatística global correspondente [18].

Por definição, um LISA deve satisfazer os seguintes requisitos: a) para cada observação fornece uma indicação da extensão do agrupamento espacial significativo de valores semelhantes em torno dessa observação; e b) a soma de LISAs para todas as observações é proporcional a um indicador global de associação espacial [17].

No caso do *Local Indicators of Spatial Association*, a soma das estatísticas locais é proporcional ao *Global Moran's I*, que corresponde à média das estatísticas locais. O LISA identifica clusters e outliers espaciais locais. Assim, o produto do valor na localização  $i$  com sua defasagem espacial e a soma ponderada dos valores nas localizações vizinhas é simplificado por Anselin [16] como  $I_i = z_i \sum_j w_{ij} z_j$ , pois a soma de todos os pesos é igual ao número total de observações ( $S_0 = n$ ).

Para interpretação do LISA é necessário analisar se os valores dos índices retornados representam significância estatística ou não. No GeoDA a ferramenta de autocorrelação espacial retorna cinco valores: i) o *Global Moran's I* (-1 a

+1); ii) o Índice esperado; iii) a variância; iv) o *z-score*; e v) o *p-value*. Os *z-scores* e os *p-values* são medidas de significância estatística que conduzem à rejeição ou não da hipótese nula de aleatoriedade espacial [19]. A hipótese nula afirma que o atributo analisado está distribuído aleatoriamente entre as feições da área de estudo, não havendo agrupamento espacial [14].

Se o *p-value* não for estatisticamente significativo, a hipótese nula não pode ser rejeitada, e possivelmente causada por processos aleatórios espacialmente. Do contrário, se o *p-value* for estatisticamente significativo a hipótese nula deve ser rejeitada. Neste caso, devem ser avaliados os valores do *z-score* [14].

Segundo Anselin [17], o *z-score* padronizado para  $I$  é calculado subtraindo o valor esperado pelo desvio padrão. Algebricamente é sintetizado por:  $Z(I) = [I - E(I)] / DP(I)$ , em que  $E(I)$  e  $DP(I)$  são, respectivamente, o valor esperado e o desvio padrão teórico de  $I$  [20]. Com o cálculo do *z-score* é possível fazer a localização de ponto na curva da distribuição normal [21].

Um *z-score* positivo indica valores semelhantes (altos ou baixos), com tendência a agrupamento [14]. Isso significa que valores altos são cercados de valores igualmente altos (*High - High*) ou valores baixos são cercados de valores igualmente baixos (*Low - Low*). Já um *z-score* negativo indica *outlier*, ou seja, valores altos são cercados por valores baixos (*High - Low*) ou valores baixos são cercados por valores altos (*Low - High*) [19]. A distribuição espacial desse conjunto de dados é mais dispersa espacialmente [14].

Se os valores no conjunto de dados tendem a se agrupar espacialmente (HH e LL), o Índice de Moran será positivo (*cluster*). Quando valores altos repelem outros valores altos e tendem a estar próximos de valores baixos, o Índice será negativo, ou seja, um *outlier* indica que uma feição tem feições vizinhas com valores diferentes [22]. Se valores de produtos cruzados positivos equilibrarem valores de produtos cruzados negativos, o Índice será próximo de zero. O numerador é normalizado pela variância, para que os valores do Índice fiquem entre -1, 0 e +1 [14].

Como resultado, o mapa de *cluster*, LISA MAP, apresenta cinco classes previamente padronizadas [10,23]: 1 - HH em vermelho (*hotspots*); 2 - LL em azul (*coldspots*); 3 - LH em azul claro; 4 - HL em rosa; 5 - em branco onde não há evidências suficiente (valor  $p > 0.05$ ) que o padrão espacial seja não-aleatório. Na análise espacial da leptospirose, os *hotspots* (HH) indicam a incidência estatisticamente alta de leptospirose, enquanto os *coldspots* (LL) implicam uma incidência estatisticamente baixa [10].

### 3. RESULTADOS

O *Global Moran's I* foi de 0,381 (pseudo *p-value* = 0,001). O pseudo *p-value* mostra que a chance de a distribuição dos dados ser aleatória é menor que 0,001. A associação espacial global foi positiva (*z-score* de 47,096), o que indica valores semelhantes, com tendência a agrupamento.

A análise da porcentagem da significância dos municípios no LISA mostra que 18,22% dos municípios foram classificados como LL. Aproximadamente 24% dos municípios (HH e LL) detectaram a existência de padrão de conglomerados espaciais (*cluster*), representando 1.329 municípios brasileiros do total de 5.570.

O período do estudo obteve 1,22% e 0,50% dos municípios com classes LH e HL, respectivamente, e a distribuição espacial desse conjunto de dados é mais dispersa espacialmente (Figura 1). Nessas duas classes, portanto, menos de 2% dos municípios foram classificados como dispersos. Mais de 70% dos municípios, no período entre 2007 e 2017, não foram significativos.

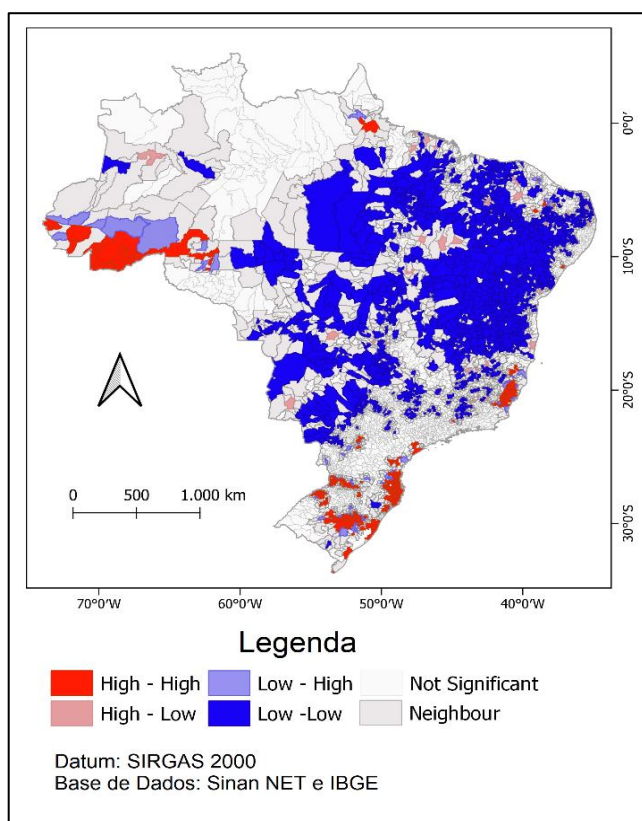


Figura 1. LISA Cluster Map da leptospirose (2007 a 2017)

O estado do Espírito Santo teve *clusters* HH em quase todo seu território. Acre e Rio Grande do Sul também apresentaram *clusters* HH, taxas altas de incidência da leptospirose cercadas por municípios vizinhos com taxas igualmente altas. Por outro lado, municípios da região central e nordeste do Brasil tiveram *clusters* LL, taxas baixas de incidência da leptospirose cercadas por municípios vizinhos com taxas igualmente baixas.

#### 4. DISCUSSÃO

Com a introdução do *Local Indicators of Spatial Association* em estudos epidemiológicos, Luenam e Puttanapong [2,10] aplicaram análises espaciais da leptospirose na Tailândia, e

da mesma forma Dhewantara [8] avaliaram o LISA no período entre 2005 e 2016 na China. Galan [9] obtiveram *clusters* da leptospirose no Brasil entre 2000 e 2015, com padrões espaciais semelhantes aos encontrados nesse artigo, mas não apresentaram os parâmetros estatísticos metodológicos, nem os valores dos Índices de Moran, inviabilizando uma análise comparativa.

O Índice de Moran mostrou significância estatística para autocorrelação espacial global (pseudo *p-value* = 0,001) e local (pseudo *p-value* = 0,05) da incidência da leptospirose no período 2007-2017 para os municípios brasileiros. Aplicando o método LISA observou-se autocorrelação em nível local e aglomerados espaciais (HH e LL), em particular as áreas com alta incidência de leptospirose (HH) como a região Sul, parte do Sudeste e sudoeste da região Norte. Aglomerados espaciais com baixas taxas de infecção (LL) foram detectados na região central do Brasil e mais de 70% dos municípios não foram significativos.

A alta frequência de municípios não significativos para este método de estatística espacial também foi observada em Galan [9], entre 2000 e 2015 no Brasil e, em Dhewantara [8] na China, entre 2005 e 2016, países com dimensões territoriais semelhantes.

Uma questão que deve ser considerada é o fato da subnotificação da doença no Brasil. Vários trabalhos trazem os vieses da subnotificação existente no setor público de saúde [24], ou das dificuldades clínicas de diagnóstico semelhantes a outras doenças tropicais negligenciadas [25,26]. Pelos motivos elucidados, pontualmente se identificou valores altos de incidência da leptospirose para alguns municípios, mas como a vizinhança não seguiu os critérios do método para significância estatística, esses municípios não foram considerados aglomerados espaciais.

#### 5. CONCLUSÕES

Este estudo explora as principais contribuições do método LISA na detecção de *clusters* espaciais da leptospirose no Brasil. Além disso, destaca as regiões Sul, Sudeste e sudoeste da região Norte, com maiores incidências de agrupamento de leptospirose. O método de análise de aglomerados espaciais atendeu ao objetivo desta pesquisa.

Sugerimos que outras variáveis ainda devam ser exploradas, principalmente para tomadas de decisões epidemiológicas. Os resultados podem sugerir a realocação de recursos na área da saúde, e melhorar a eficácia da prevenção e hospitalização para a redução da taxa de leptospirose.

O presente trabalho foi realizado com apoio da CAPES, Código de Financiamento 001.

#### 6. REFERÊNCIAS

[1] World Health Organization. *Human leptospirosis: guidance for diagnosis, surveillance and control*. WHO, pp. 109, Malta, 2003.

- [2] A. Luenam, and N. Puttanapong. Spatial and statistical analysis of leptospirosis in Thailand from 2013 to 2015. *Geospatial Health*. v.14(1), pp.121–7, 2019.
- [3] L. R. Londe, R. S. Conceição, T. Bernardes, and M. C. A. Dias. Flood-related leptospirosis outbreaks in Brazil: perspectives for a joint monitoring by health services and disaster monitoring centers. *Nat Hazards*, v.84(2), pp.1419–35, 2016.
- [4] A. N. Marteli, L. V. Genro, D. Diament, and L. A. Guasselli. Análise espacial da leptospirose no Brasil. *Saúde em Debate*. 2020.
- [5] O. S. Baquero, and G. Machado. Spatiotemporal dynamics and risk factors for human Leptospirosis in Brazil. *Sci Rep*. v. 8(1), 2018.
- [6] J. D. Gutiérrez, and et al. Environmental and socioeconomic determinants of leptospirosis incidence in Colombia. *Cad Saude Publica*. v.35(3), 2019.
- [7] G. G. Pedra. Predicting environmental risk of transmission of leptospirosis. University of Liverpool; 2019.
- [8] P. W. Dhewantara, and et al. Geographical and temporal distribution of the residual clusters of human leptospirosis in China, 2005–2016. *Sci Rep*. v.8(1), 2018.
- [9] D. I. Galan, and et al. Epidemiology of human leptospirosis in urban and rural areas of Brazil, 2000–2015. *PLoS One*. Mar v.16(3), 2021.
- [10] A. Luenam, N. Puttanapong. Modelling and analyzing spatial clusters of leptospirosis based on satellite-generated measurements of environmental factors in Thailand during 2013-2015. *Geospatial Health*, v. 15(2), 2020, 217–24.
- [11] Ministério da Saúde (Brasil). Ministério da Saúde (Brasil). Departamento de Informática do Sistema Único de Saúde. Doenças e Agravos de Notificação - de 2007 em diante (SINAN) Departamento de Informática do Sistema ico de Saúde – DATASUS. *Leptospirose: Casos confirmados por município de residência de 2007 – 2017*. Brasília - DF, 2019.
- [12] Ministério da Saúde (Brasil). Sistema de Informação de Agravos de Notificação. População residente por Município e Ano – DATASUS. Brasília - DF, 2019.
- [13] P.A.P. Moran. The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society: Series B (Methodological)*, Volume 10(2), 1948.
- [14] Esri. Documentação ArcGis Pro, Califórnia. 2022 [cited 2022 Jan 31]. How Spatial Autocorrelation [Global Moran's I] works. Disponível em: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm>
- [15] Anselin L. geodacenter. 2020. Global Spatial Autocorrelation. Disponível em: [https://geodacenter.github.io/workbook/5a\\_global\\_auto/lab5a.html](https://geodacenter.github.io/workbook/5a_global_auto/lab5a.html)
- [16] Anselin L. geodacenter. 2020. Contiguity-Based Spatial Weights. Disponível em: [http://geodacenter.github.io/workbook/4a\\_contig\\_weights/lab4a.html](http://geodacenter.github.io/workbook/4a_contig_weights/lab4a.html)
- [17] L. Anselin. Local Indicators of Spatial Association [LISA]. *Geogr Anal*. 1995; 27:23. Disponível em: <https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1538-4632.1995.tb00338.x>
- [18] L. Anselin. Geodacenter. Local Spatial Autocorrelation LISA and Local Moran. 2020. Disponível em: [https://geodacenter.github.io/workbook/6a\\_local\\_auto/lab6a.html](https://geodacenter.github.io/workbook/6a_local_auto/lab6a.html)
- [19] Documentação ArcGis Pro - Esri. Cluster and Outlier Analysis (Anselin Local Moran's I). Spatial Statistics. Califórnia. 2022 Disponível em: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/cluster-and-outlier-analysis-anselin-local-moran-s.htm>
- [20] E. Almeida. Análise exploratória de dados espaciais. In: Almeida E, editor. *Econometria espacial aplicada*. Campinas: Alínea, 2012. Disponível em: [http://www.usp.br/nereus/wp-content/uploads/Almeida\\_cap\\_4.pdf](http://www.usp.br/nereus/wp-content/uploads/Almeida_cap_4.pdf)
- [21] J. K. Yamamoto. Como Calcular o Índice Moran – Autocorrelação Espacial. 2022. Disponível em: <https://geokrigagem.com.br/como-calcular-o-indice-moran-autocorrelacao-espacial/>
- [22] Documentação ArcGis Pro (Esri). How Cluster and Outlier Analysis [Anselin Local Moran's I] works. Califórnia. 2022.
- [23] Anselin L. Exploring Spatial Data with GeoDaTM: A Workbook. 2005.
- [24] Ministério da Saúde. *Guia de Vigilância em Saúde*. 3rd ed. Coordenação-Geral de Desenvolvimento da Epidemiologia em Serviços. Brasília - DF: Ministério da Saúde, pp.740, 2019.
- [25] F. Costa, J. E. Hagan, J. Calcagno, M. Kane, P. Torgerson, M. S. Martinez-Silveira, et al. Global Morbidity and Mortality of Leptospirosis: A Systematic Review. *PLOS Neglected Tropical Diseases*, v. 9(9), 2015.
- [26] M. Picardeau. Leptospirosis: Updating the Global Picture of an Emerging Neglected Disease. *PLOS Neglected Tropical Diseases*, v. 9(9), 2015.