



UNIVERSIDADE FEDERAL DO PARÁ  
CAMPUS UNIVERSITÁRIO DE CASTANHAL  
FACULDADE DE COMPUTAÇÃO  
BACHARELADO EM ENGENHARIA DE COMPUTAÇÃO

JULIO MOREIRA SOARES NETO

**EPIDEMIOLOGIA ESPACIAL DA HANSENÍASE NO MUNICÍPIO DE  
PARAGOMINAS, PARÁ**

CASTANHAL-PA

2019

JULIO MOREIRA SOARES NETO

**EPIDEMIOLOGIA ESPACIAL DA HANSENÍASE NO MUNICÍPIO DE  
PARAGOMINAS, PARÁ**

Trabalho de Conclusão de Curso, apresentado como requisito parcial para obtenção de grau de Bacharel em Engenharia de Computação, pela Universidade Federal do Pará.

Orientador: Prof. Dr. Josafá Gonçalves Barreto.

Coorientador: Prof. Dr. Jose Jailton Henrique Ferreira Junior

CASTANHAL-PA

2019

JULIO MOREIRA SOARES NETO

**EPIDEMIOLOGIA ESPACIAL DA HANSENÍASE NO MUNICÍPIO DE  
PARAGOMINAS, PARÁ**

Trabalho de Conclusão de Curso,  
apresentado como requisito parcial para  
obtenção de grau de Bacharel em Engenharia  
de Computação, pela Universidade Federal  
do Pará.

Data de aprovação: \_\_\_/\_\_\_/\_\_\_

**Banca Examinadora:**

---

Prof. Dr. Josafá Gonçalves Barreto

Orientador - Faculdade de Educação Física – UFPA - Castanhal

---

Prof. Dr. Jose Jailton Henrique Ferreira Junior

Coorientador - Faculdade de Computação (FACOMP) – UFPA - Castanhal

---

Prof. Dr. Tássio Costa de Carvalho

Membro Interno - Faculdade de Computação (FACOMP) – UFPA - Castanhal

---

Prof. Dr. Marcos César da Rocha Seruffo

Membro Externo – Faculdade de engenharia da computação e  
telecomunicações (FCT) – UFPA - Belém

A Samia Laiane.

## **AGRADECIMENTOS**

Ao Prof. Dr. Josafá Gonçalves Barreto, pela orientação durante o TCC e na bolsa de iniciação científica.

Ao Prof. Dr. Jose Jailton Henrique Ferreira Junior, pela atenção e orientação na elaboração do TCC.

Ao Prof. Dr. Marcos Seruffo, pela paciência e presença nos meus “primeiros passos” na iniciação científica.

Aos colegas do Laboratório de Epidemiologia Espacial (LabEE): Junior e Pedro, pela disposição e paciência para me ajudar a resolver os problemas encontrados nos meus primeiros dias como bolsistas.

Aos colegas das viagens de trabalho de campo, com quem aprendi muito e sem o trabalho realizado por eles, esse estudo não poderia ser feito.

A todos os professores da FACOMP pela instrução, paciência e inspiração ao longo do curso.

Aos colegas do Laboratório de Desenvolvimento de Sistemas (LADES): Arnaldo, Carlos, Igor, Heitor, Hygor e Kennedy, pela companhia e apoio durante o tempo em que fui bolsista.

Aos amigos e colegas que me ajudaram durante as análises de alguns resultados e na discussão de alguns trechos da monografia e por me ajudar a manter o foco no estudo: Hygor, Victor, Francisco, Douglas e Sâmia.

A minha família pelo incentivo a minha educação.

Ao Programa Institucional de Bolsas de Iniciação Científica – PIBIC pela concessão da bolsa de Iniciação Científica PIBIC UFPA interior.

“Foi dito que 'a verdade libertará os homens'. A verdade por si só nunca fez ninguém livre. É apenas a dúvida que trará a emancipação mental”.

(Anton LaVey)

## RESUMO

A hanseníase é uma doença infectocontagiosa crônica causada pelo *Mycobacterium leprae*, que afeta principalmente a pele e nervos periféricos, capaz de causar incapacidades físicas e deformidades quando não diagnosticada e tratada precocemente. Trinta por cento dos municípios paraenses são hiperendêmicos para hanseníase, como é o caso de Paragominas que possui uma taxa de detecção de 46,98 /100.000 hab em 2017. Para identificar a distribuição espacial dos casos de hanseníase residentes em Paragominas, os casos foram mapeados utilizando aparelho de GPS com os endereços fornecidos pela Secretaria Municipal de Saúde. Os dados coletados foram processados no *software* QGIS para elaboração dos mapas de distribuição de casos, taxa de detecção por setores censitários e mapa de calor. Para a análise estatística foram aplicados os testes I de Moran, local e global, varredura de Kulldorff e análise espaço-temporal de Knox para detecção de *clusters* utilizando-se os programas GeoDa, *Clusterseer* e PPA. Um total de 1047 casos novos foram notificados entre 2004 e 2010, 791 residiam na zona urbana e 91% deles foram mapeados. Foi observada uma média de nove casos por setor censitário (mínimo 0 e máximo 33); 76 (95%) dos 80 setores censitários urbanos foram classificados como hiperendêmicos. As análises espaciais mostraram a formação de *cluster* estatisticamente significativa ( $p= 0,001$  para I de Moran, Kulldorff e Knox). O teste de Knox mostrou que 83% dos casos tem conexão espaço-temporal com outro caso em um raio de 100 metros no período de 3 anos de detecção. A distribuição espacial da hanseníase em Paragominas é heterogênea, com formação de *clusters*. Os resultados mostram que o uso dos sistemas de informação geográfica ajuda a entender a dinâmica da transmissão da doença no município, além de contribuir no planejamento das ações de controle. Adicionalmente, realizou-se uma descrição detalhada dos métodos de estatística espacial utilizados neste estudo.

**Palavras-chave:** Epidemiologia Espacial. Hanseníase. Sistemas de Informações Geográficas.

## ABSTRACT

Leprosy is a chronic infectious disease caused by *Mycobacterium leprae*, which mainly affects the skin and peripheral nerves, it can cause physical disabilities and deformities when isn't diagnosed and treated early. Thirty percent of the municipalities of Pará are hyperendemic for leprosy, as is the case of Paragominas, which has a detection rate of 46.98 / 100,000 inhabitants in 2017. To identify the spatial distribution of leprosy cases resident in Paragominas, the cases were mapped using a GPS device with the addresses provided by the Municipal Health Department. The data collected were processed in the QGIS software to elaborate case distribution maps, detection rate by census tracts and heat map. For statistical analysis, the local and global Moran I tests were applied, as well as Kulldorff and Knox scanning for clusters detection using the programs GeoDa, ClusterSeer and PPA. A total of 1047 new cases were reported between 2004 and 2010, 791 resided in the urban area and 91% were mapped. An average of nine cases per census tract was observed (minimum 0 and maximum 33); 76 (95%) of the 80 urban census tracts were classified as hyperendemic. Spatial analyzes showed statistically significant cluster formation ( $p = 0.001$  for I by Moran, Kulldorff and Knox). The Knox test showed 83% of the cases have a space-time connection with another case within 100 meters of radius in the period of 3 years of detection. The spatial distribution of leprosy in Paragominas is heterogeneous, resulting in clusters of high transmission. The results help to understand the dynamics of disease transmission in the municipality, besides contributing to the planning of control actions. In addition, a detailed description of the spatial statistics methods used in this study was made.

**Keywords:** Spatial Epidemiology. Leprosy. Geographic Information Systems.

## LISTA DE ILUSTRAÇÕES

|   |    |
|---|----|
| Figura 1 – Distribuição de novos casos de hanseníase em 2017 .....  | 13 |
| Figura 2 - Taxa de detecção da hanseníase no Brasil, 2016. ....   | 14 |
| Figura 3 - Taxa de detecção anual da hanseníase nos municípios do Pará, 2017. ....  | 15 |
| Figura 4 - Distribuição espacial das mortes por cólera em Londres, 1854.....  | 16 |
| Figura 5 - Estimativa de <i>Kernel</i> . ....   | 24 |
| Figura 6 - Representação gráfica da varredura espacial.....   | 26 |
| Figura 7 - Distribuição de casos de Hanseníase na zona urbana de Paragominas, Pará, 2004-2010. ....   | 32 |
| Figura 8 - Mapa de calor dos casos de hanseníase na zona urbana de Paragominas, Pará, 2004-2010.....  | 33 |
| Figura 9 - Taxa de detecção de casos novos de hanseníase, na zona urbana de Paragominas, Pará, 2004-2010.....                                       | 34 |
| Figura 10 - Método de Kulldorff para detecção de <i>clusters</i> de casos de hanseníase, na zona urbana de Paragominas, Pará, 2004-2010.....        | 35 |
| Figura 11 - Gráfico de dispersão de I de Moran.....   | 37 |
| Figura 12 - Mapa de significância LISA.....   | 38 |
| Figura 13 - <i>Cluster LISA</i> de casos de hanseníase, na zona urbana de Paragominas, Pará, 2004-2010. ....  | 39 |
| Figura 14 - Método de Knox para detecção de links espaço-temporais, no <i>cluster</i> principal da zona urbana de Paragominas, Pará, 2004-2010..... | 41 |
| Figura 15 - Distribuição de renda, população e taxa de moradores por residência em Paragominas, Pará, 2010.....                                     | 42 |

## LISTA DE TABELAS

|   |    |
|---|----|
| Tabela 1 - Autocorrelação espacial global de Moran para a taxa de detecção anual. ... | 36 |
| Tabela 2 - Análise espaço temporal de Knox para casos de hanseníase. ....             | 40 |

## LISTA DE SIGLAS

|          |   |
|----------|---|
| CEP-     | Comitê de ética em pesquisa/Instituto de ciências da saúde/Universidade |
| ICS/ufpa | Federal do Pará   |
| GIS      | Geographic information system   |
| GPS      | Global positioning system   |
| IBGE     | Instituto brasileiro de geografia e estatística                         |
| LISA     | Local indicator of spatial association                                  |
| OMS      | Organização mundial da saúde  |
| PA       | Pará  |
| SIG      | Sistemas de informação geográfica                                       |
| SINAN    | Sistema de informação de agravos de notificação                         |
| SIRGAS   | Sistema de referência geocêntrico para as américas                      |
| WHO      | World health organization   |

## SUMÁRIO

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>INTRODUÇÃO.....</b>   | <b>12</b> |
| 1.1      | Aspectos gerais da hanseníase.....                             | 12        |
| 1.2      | Epidemiologia .....  | 12        |
| 1.3      | Epidemiologia Espacial .....                                   | 15        |
| 1.4      | Sistemas de informação geográfica.....                         | 17        |
| 1.5      | Objetivos .....  | 18        |
| 1.5.1    | Objetivo geral.....  | 18        |
| 1.5.2    | Objetivo específico.....                                       | 19        |
| <b>2</b> | <b>METODOLOGIA.....</b>  | <b>20</b> |
| 2.1      | Preparação dos mapas.....                                      | 20        |
| 2.2      | Software Qgis.....   | 20        |
| 2.3      | Software GeoDa .....   | 21        |
| 2.4      | Software Point Pattern Analysis (PPA) .....                    | 21        |
| 2.5      | Software ClusterSeer 2.5.....                                  | 22        |
| <b>3</b> | <b>ESTATÍSTICA ESPACIAL .....</b>                              | <b>23</b> |
| 3.1      | Taxa de detecção de casos novos .....                          | 23        |
| 3.2      | Densidade de <i>kernel</i> .....                               | 23        |
| 3.3      | Simulações de Monte Carlo .....                                | 25        |
| 3.4      | Método de detecção de <i>cluster</i> de Kulldorff.....         | 25        |
| 3.5      | Indicador de autocorrelação espacial.....                      | 28        |
| 3.5.1    | Global .....   | 28        |
| 3.5.2    | Local .....  | 29        |
| 3.6      | Método de detecção de interações espaço-temporais de Knox..... | 30        |
| <b>4</b> | <b>RESULTADOS .....</b>  | <b>32</b> |
| <b>5</b> | <b>DISCUSSÃO.....</b>  | <b>43</b> |
| <b>6</b> | <b>CONCLUSÃO.....</b>  | <b>46</b> |
|          | REFERÊNCIAS.....   | 47        |

# 1 INTRODUÇÃO

## 1.1 Aspectos gerais da hanseníase

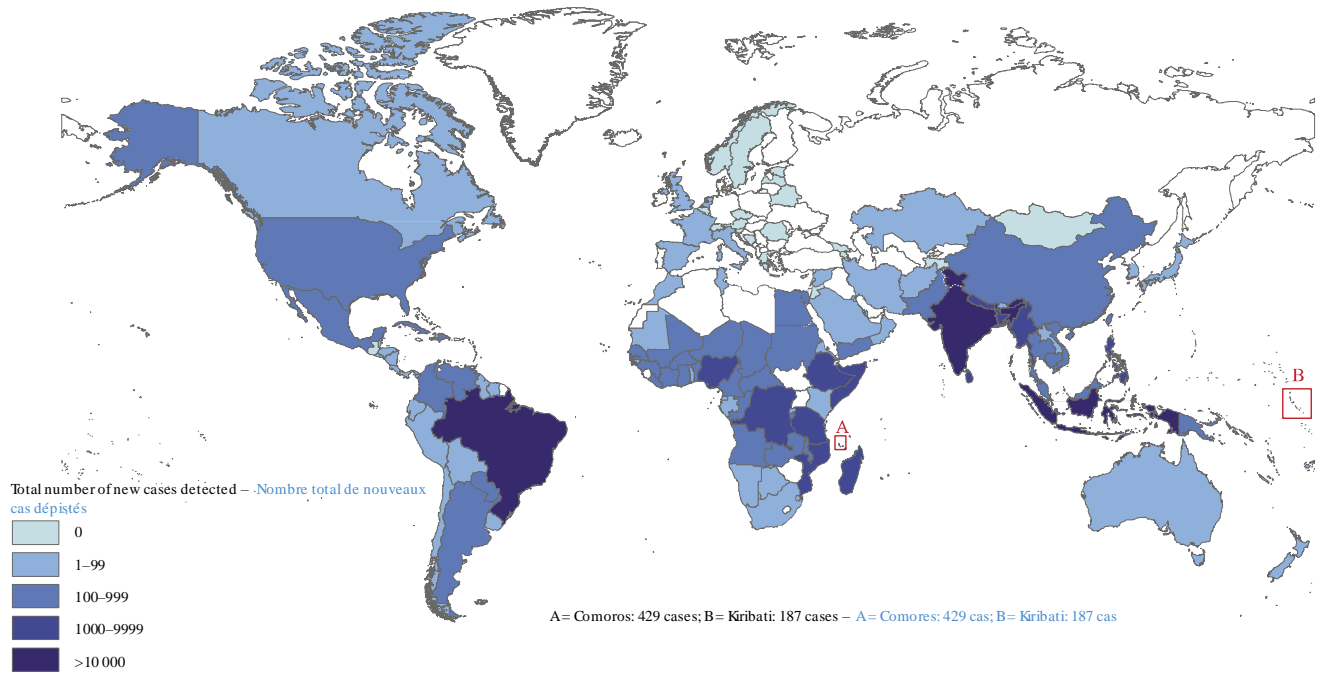
A hanseníase é uma doença infectocontagiosa crônica causada pelo *Mycobacterium leprae*, bacilo identificado como agente etiológico da doença pelo médico norueguês Gerhard Armauer Hansen (1874). A doença afeta principalmente a pele e nervos periféricos, capaz de causar incapacidades físicas e deformidades quando não diagnosticada e tratada precocemente. A hanseníase varia de uma infecção paucibacilar com um número reduzido de lesões, a uma doença multibacilar com diversas lesões na pele, nervos periféricos, nos olhos e vias aéreas. Possui um longo período de incubação, com média de 3 a 5 anos, podendo se estender por décadas, o ambiente domiciliar é apontado como meio facilitador no processo de transmissão, prolongando o contato entre familiares saudáveis e infectados, aumentando de cinco a dez vezes as chances de infecção (WALKER; LOCKWOOD, 2006). Não existe exame laboratorial capaz de diagnosticar todas as formas clínicas da hanseníase e o diagnóstico é feito pelo exame clínico dermatoneurológico (SECRETARIA DE VIGILÂNCIA EM SAÚDE, 2018).

Os pacientes com o tipo multibacilar da hanseníase sem tratamento podem ser as principais fontes de infecção e os seus contatos intradomiciliares são o principal grupo de risco, e mesmo os contatos de pacientes paucibacilares tem um risco maior de infecção em comparação com a população em geral (RICHARDUS, 2004). Mesmo em regiões de alta endemicidade somente uma pequena parte da população exposta a doença, estima-se em 10%, desenvolve a doença (CREE; SMITH, 1998).

## 1.2 Epidemiologia

Em 2017, 210.671 casos novos foram reportados à Organização Mundial da Saúde (OMS) por 150 países e a taxa de detecção global de novos casos foi de 2,77/100.000 habitantes (WHO, 2018). O Brasil registrou 26.875 casos em 2017 e aparece em destaque no mapa da Figura 1.

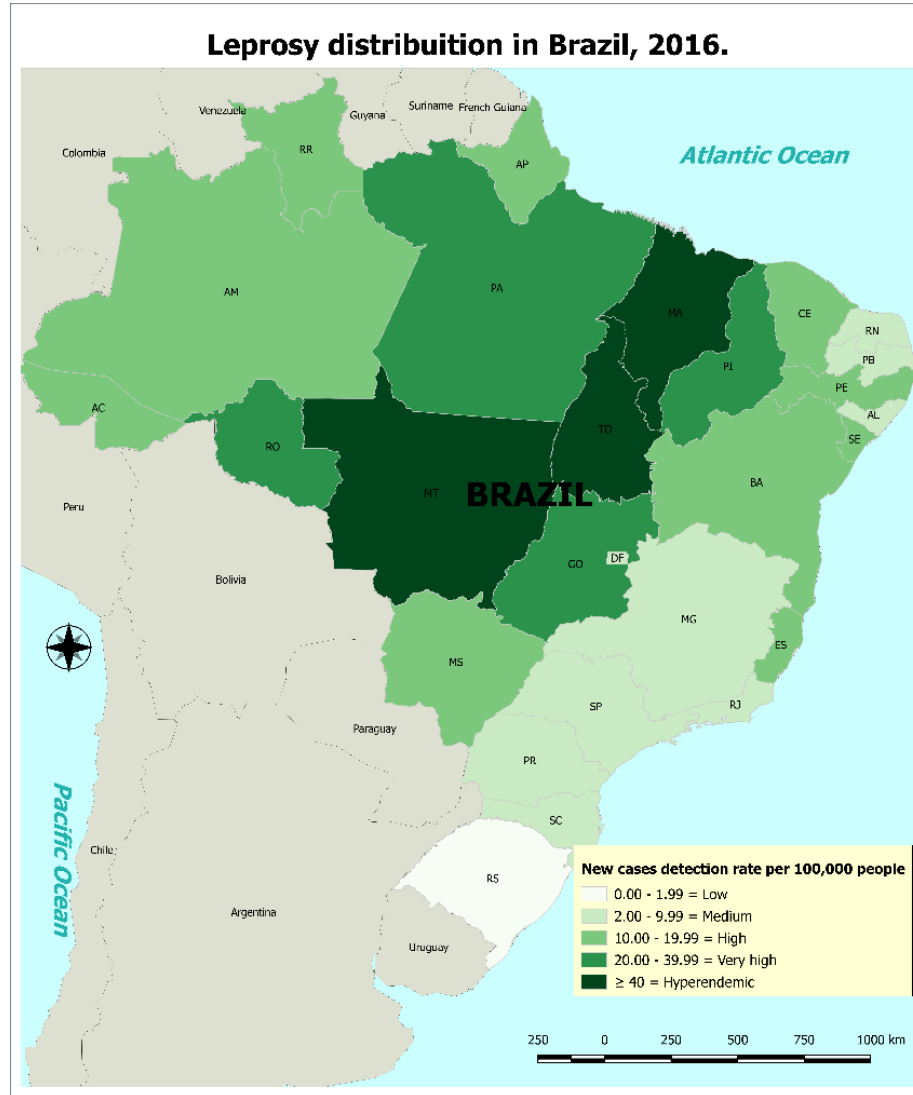
**Figura 1** – Distribuição de novos casos de hanseníase em 2017



Fonte: (WHO, 2018)

A distribuição da hanseníase no Brasil não apresenta uniformidade, existindo aglomerações de casos em algumas regiões do país (PENNA et al., 2009). A Figura 2 mostra a taxa de detecção da hanseníase no Brasil em 2016.

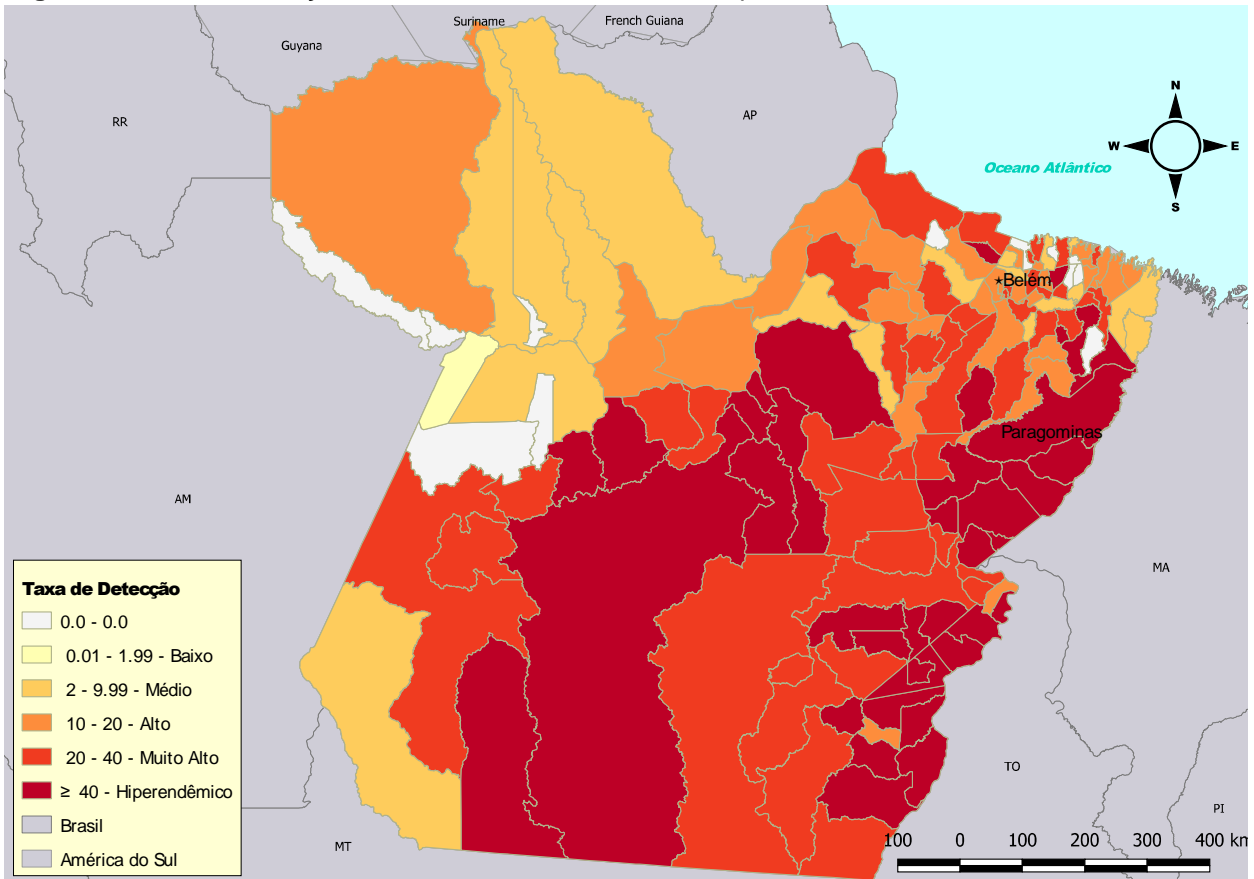
**Figura 2 - Taxa de detecção da hanseníase no Brasil, 2016.**



Fonte: LabEE (2017).

De 2001 a 2017 mais de 70.000 casos de hanseníase foram diagnosticados no Pará, com um coeficiente de detecção anual em 2017 de 31,05 para 100.000 habitantes e taxa de prevalência de 3,2 para 10 mil habitantes (BRASIL, 2018), a doença permanece como um grave problema de saúde pública neste estado. Em 2017 trinta e sete dos 144 municípios paraenses são hiperendêmicos como mostra a Figura 3.

**Figura 3 - Taxa de detecção anual da hanseníase nos municípios do Pará, 2017.**



**Fonte:** Autor (2018).

Paragominas, cidade localizada a 300km da capital, com população de 110.026 habitantes apresentou, em 2017, uma taxa de detecção de 46,98 para cada 100.000 habitantes e taxa de prevalência de 6,45 para 10 mil habitantes e também é classificada como hiperendêmica para hanseníase (BRASIL, 2018). O entendimento da distribuição espacial da hanseníase no Pará contribui para o aumento da detecção precoce de casos novos, guiando estratégias de busca ativa em áreas de maior risco de transmissão da doença (BARRETO et al., 2015).

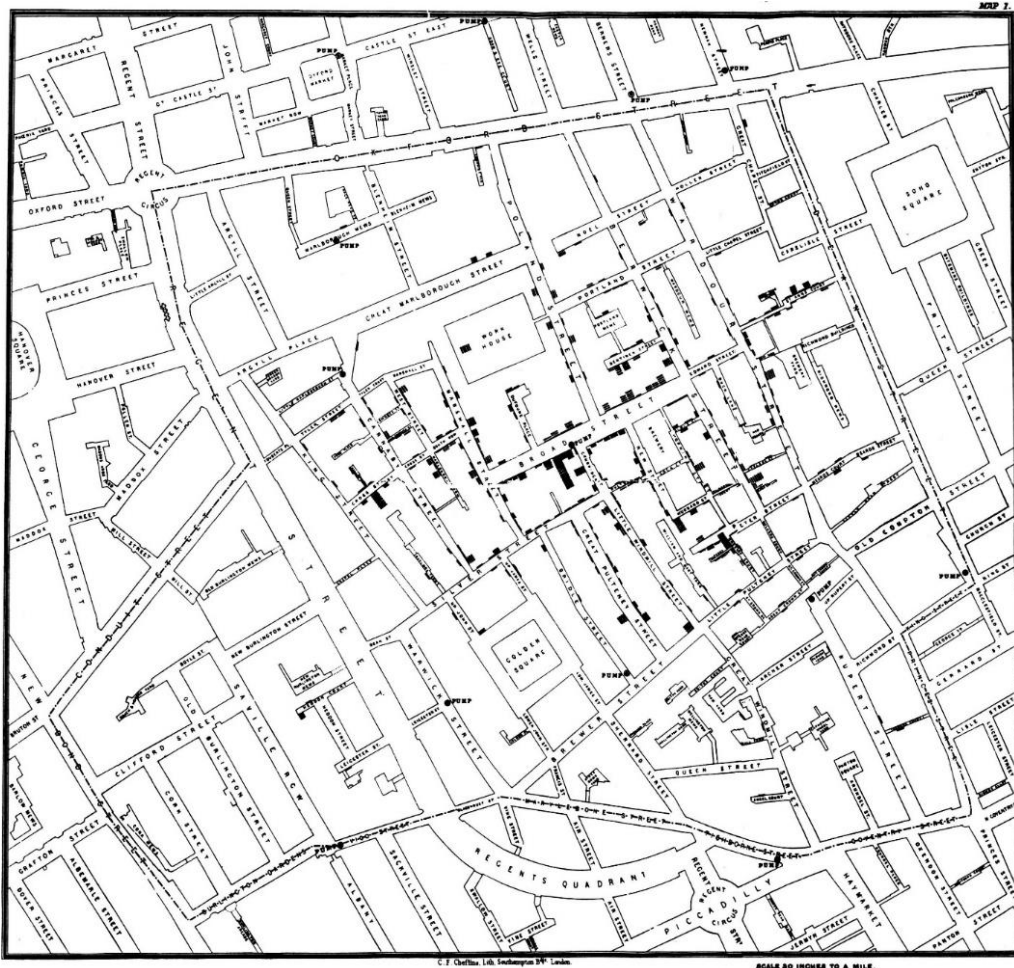
### 1.3 Epidemiologia Espacial

Epidemiologia espacial trata da distribuição geográfica de dados relacionados a saúde, fatores de risco demográficos, ambientais, comportamentais, socioeconômicos, genéticos e infecciosos (ELLIOTT; WARTENBERG, 2004). Na análise espacial

epidemiológica, o estudo será guiado pela natureza e estilo dos métodos, que consiste em uma hipótese sobre os dados observados. É conveniente classificar as hipóteses em três grandes classes: Mapeamento da distribuição de casos, análise ecológica e agrupamento (*cluster*) de casos (LAWSON, 2013).

Os primeiros trabalhos de epidemiologia espacial datam do século XIX, com mapas usados para entender a propagação e surtos de doenças infecciosas. O médico John Snow, foi o pioneiro na epidemiologia espacial com a investigação espacial das mortes causadas por cólera, ele mapeou os óbitos na cidade de Londres em 1854, identificando uma concentração de óbitos entre os residentes que utilizavam a água proveniente de uma bomba localizada na rua Broad (Figura 4). A bomba foi desativada, e o número de casos novos de cólera diminuiu drasticamente (SNOW, 1967).

**Figura 4** - Distribuição espacial das mortes por cólera em Londres, 1854.



Fonte: Mode of communication of cholera (SNOW, 1967).

O mapeamento preciso e rápido, com várias técnicas de exploração, análise de grandes volumes dados georreferenciados e com diversas variáveis é possível na epidemiologia graças aos avanços na capacidade de processamento dos computadores pessoais, o desenvolvimento do sistema de posicionamento global (GPS), o aumento no número de ferramentas dos sistemas de informação geográfica (SIG) e a disponibilização desses dados espaciais, facilitado para que os pesquisadores possam entender a dinâmica dos dados investigados.

#### **1.4 Sistemas de informação geográfica**

Bonham-Carter (1994) define os SIG separando e explicando o que cada palavra significa:

- a) sistema composto por componentes com diversas funções em uma interface para manipulação, visualização, análise, investigação e modelagem;
- b) informacional porque os dados são organizados para produzir conhecimentos úteis;
- c) geográfico, pois a localização dos dados é conhecida ou pode ser calculada em termos de latitude e longitude.

De acordo com Waller e Gotway (2004), mesmo que os sistemas variem em interface e funcionalidade, todos os SIGs possuem três funções básicas para analisar os dados espaciais:

- a) investigação espacial permite a visualização dos dados e seus atributos em suas respectivas localizações espaciais, com isso é possível classificar, agrupar e analisar os dados, em relação a distância espacial entre um evento e outro;
- b) disposição de camadas é a função de sobrepor camadas de diferentes conjuntos de dados, a fim de montar um mapa empilhando e ligando outros mapas com informações diferentes (ex. setores do censo, divisão política, demografia, estradas, rios, etc.) da mesma área geográfica de estudo, respondendo questões que não poderiam ser abordadas analisando os dados de forma separada;

- c) buffer é a criação de áreas ao redor de pontos, linhas e polígonos, é usada para ligar os casos, para definir a densidade de eventos, ou agrupar os eventos dentro de uma área de influência definida.

Os SIG facilitaram a pesquisa para dados dependentes espacialmente e facilitaram a análise de dados que são espacialmente heterogêneos. No trabalho de Anselin e Getis (1992) é apontado que a disseminação dos resultados da análise de dados de forma empírica era limitada, e que com a aplicação dos SIGs o problema foi praticamente eliminado. Além da exibição de dados espacialmente, os SIGs possuem várias ferramentas de análise espacial e estatística espaço-temporais, que se baseiam em simulações para verificar se nos dados observados estão simplesmente distribuídos de forma aleatória ou se algo de especial (aglomerações, casos isolados ou correlações) está acontecendo para os dados estejam distribuídos da forma observada, antes das simulações e do avanço na complexidade das ferramentas dos SIGs, as distribuições eram aproximadas de forma analítica, usando modelos matemáticos complexos para representar os dados, esses modelos podem ser muito difíceis ou impossíveis de serem aplicados dependendo dos dados, principalmente se for em grande volume de casos, variáveis ou subdivisões espaciais onde os dados estão distribuídos (KULLDORFF, 1997).

A OMS estimula o uso de SIGs no monitoramento e análise de dados epidemiológicos da hanseníase a nível global, regional e sub-regional para identificação de tendências, correlações e aglomerações, além de ser uma importante ferramenta de gerenciamento das estratégias e ações para a eliminação da doença (OMS, 2016).

## **1.5 Objetivos**

### **1.5.1 Objetivo geral**

Identificar a distribuição espacial e temporal de casos de hanseníase no período de 2004 a 2010 no município de Paragominas-PA.

### **1.5.2 Objetivo específico**

- a) Demonstrar as funções, utilidades e aplicações de um SIG para o mapeamento de casos registrados de hanseníase em Paragominas.
- b) Descrever as fórmulas e aplicações de análises estatísticas espaciais e temporais.
- c) Identificar o padrão de distribuição espacial e temporal dos casos de hanseníase no período estudado.

## **2 METODOLOGIA**

Este estudo está em conformidade com a Declaração de Helsinki e foi aprovado pelo Comitê de Ética em Pesquisa do Instituto de Ciências da Saúde da Universidade Federal do Pará (protocolo número 197/07 CEP-ICS/UFGPA).

### **2.1 Preparação dos mapas**

Este estudo foi iniciado com coleta de dados espaciais por meio de trabalho de campo no período 01 a 06 de novembro de 2010 com uma equipe de pesquisadores composta por médicos, fisioterapeutas e técnicos de saúde visitaram as famílias listadas no banco de dados para avaliar casos e examinar os contatos intradomiciliares. Os dados sobre os casos de hanseníase de 2004 a 2010 notificados no município de Paragominas, Pará foram coletados da base do Sistema de Informação de Agravos de Notificação (SINAN) e fornecidos pela Secretaria Municipal de Saúde.

Durante as visitas a equipe fazia a geolocalização dos pacientes, usando o aparelho de recepção de GPS para a criação de uma camada de pontos com as informações dos casos de Hanseníase usando a projeção do Sistema de Referência Geocêntrico para as Américas (SIRGAS 2000). Os casos e os dados são listados em forma de tabela para importação e união com os dados do Instituto Brasileiro de Geografia e Estatística (IBGE) e SINAN, a união é feita usando o número de notificação como chave primária para identificar cada caso.

Para elaboração das análises espaciais da doença, as informações dos pacientes foram agregadas por setores censitários do censo de 2010 do IBGE. Paragominas é dividida, de acordo com o censo de 2010, em 119 setores censitários, 80 setores compõem a zona urbana do município, que é o foco desse trabalho.

### **2.2 Software Qgis**

O Qgis é um SIG de código aberto usado para visualização da distribuição geográfica de dados dispostos em camadas de pontos, linhas ou polígonos. O sistema

possui várias ferramentas de análises espaciais e temporais, além de uma ferramenta para renderização e exportação de mapas para vários formatos. Além das ferramentas nativas, o software possui uma biblioteca online de plugins feitos por outros usuários que podem ser adicionados e acessados direto da janela de ferramentas do Qgis. Outras análises que foram feitas usando pacotes de *software* de estatística foram importadas para o Qgis a fim de usar a ferramenta de renderização de mapas.

### **2.3 Software GeoDa**

GeoDa é um *software* livre de graça desenvolvido para servir como uma ferramenta de análise espacial, modelagem de padrões espaciais e a possibilidade de trabalhar com uma grande variedade de formatos de dados espaciais. Entre as ferramentas espaciais do programa temos autocorrelação espacial local desenvolvida em conjunto com o Dr. Luc Anselin (2003, 2006), com opções de investigação de *clusters* espaciais com variáveis simples e múltiplas variáveis, a matriz de contiguidade, que define os vizinhos próximos é efetuada antes da estatística. Além da modelagem de mapas também é possível visualizar os dados espaciais em gráficos de dispersão, histogramas e gráficos de média.

### **2.4 Software Point Pattern Analysis (PPA)**

O PPA é um programa de graça com código fonte livre de análise espacial que opera em dados univariáveis. O *software* é usado nesse estudo para análise de autocorrelação global no método I de Moran, mas possui outros métodos como os de K de Ripley's, análise de vizinho mais próximo e autocorrelação de C de Geary's (ALDSTADT; GETIS, 2002). O programa é desenvolvido em C executável em terminal DOS ou terminal Linux. Para o uso do *software* os dados investigados devem ser tratados e distribuídos em um arquivo ASCII, possuindo somente números divididos em três colunas de dados x e y representando as dimensões espaciais e z representado a variável, para então calcular a matriz espacial de pesos, que define a vizinhança de cada um dos pontos, e então efetuar as análises estatísticas.

## 2.5 Software ClusterSeer 2.5

O ClusterSeer é um programa proprietário desenvolvido pela BioMedware, a compra da licença não dá acesso a futuras atualizações do *software* que atualmente encontra-se na versão 2.5. O programa possui 15 métodos de identificação de *clusters* espaciais, temporal e espaço-temporal. Além de identificar as aglomerações o *software* usa métodos para calcular a significância estatística dos resultados obtidos, no total o ClusterSeer oferece 24 métodos de análises estatísticas. O programa aceita vários tipos de formato de dados espaciais, sendo necessário fazer poucos ajustes nos dados para que o *software* possa começar a calcular as análises que possuem diversas opções para definir a matriz de contiguidade e distribuição temporal dos dados. No final da análise temos os mapas com os clusters, histogramas para analisar a distribuição e um arquivo de texto mostrando os dados da análise.

### 3 ESTATÍSTICA ESPACIAL

#### 3.1 Taxa de detecção de casos novos

A taxa de detecção geral de hanseníase é usada para classificar o grau de endemia da doença usando o número de casos novos diagnosticados, por 100 mil habitantes, na população de determinado espaço geográfico, durante o período considerado, matematicamente representado por:

$$\frac{\left(\frac{\text{Casos}}{\text{População}}\right) * 100000(\text{habitantes})}{\text{período (anos)}} \quad (1)$$

O valor da taxa é usado pelo Ministério da Saúde do Brasil para classificar a endemicidade da área estudada em: baixo (0,01 – 1,99), médio (2 – 9,99), alto (10 – 19,99), muito alto (20 – 39,99) e Hiperendêmico ( $\geq 40$ ).

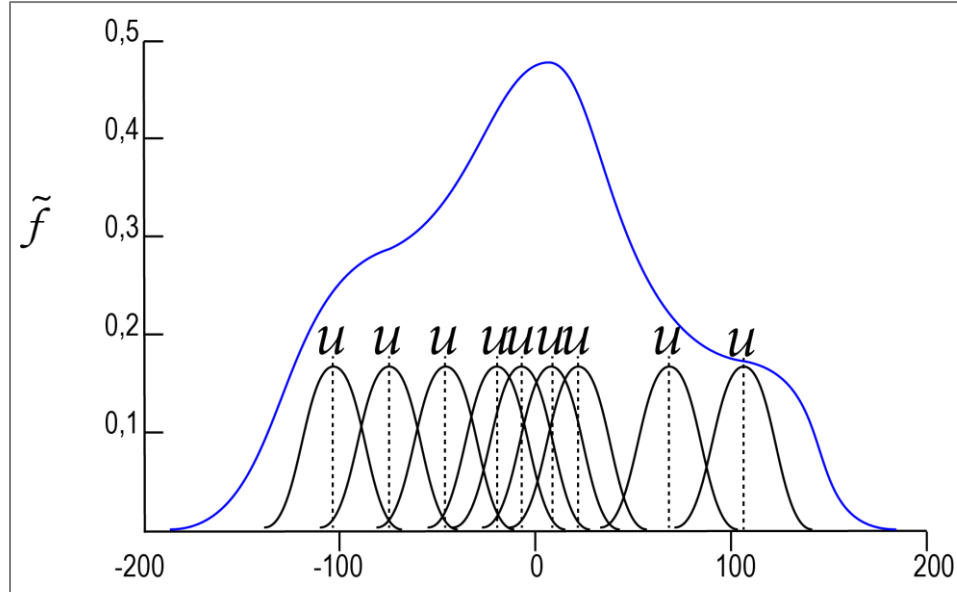
#### 3.2 Densidade de *kernel*

O objetivo dessa técnica é mostrar a concentração de eventos, originalmente representado em vetores, interpolando os pontos vizinhos dentro de um raio de influência definido, transformando os dados vetoriais em uma imagem rasterizada onde cada pixel recebe um valor. Os valores dos pixels podem ser visualizados como elevações ou gradiente de cores conhecido como mapa de calor.

Devido à natureza bidimensional da distribuição dos pontos, é necessário usar o produto de dois *kernels* unidimensionais para as coordenadas  $u_0$  e  $v_0$  de cada caso

$$\tilde{f}(u_0, v_0) = \frac{1}{Nb_u b_v} \sum_{i=1}^N \left\{ kern\left(\frac{u_0 - u_i}{b_u}\right) kern\left(\frac{v_0 - v_i}{b_v}\right) \right\}, \quad (2)$$

onde *kern* é a função de *kernel* escolhida e  $b$  é o parâmetro de suavização, ( $\tilde{f}$ ) é a estimativa de *kernel*,  $N$  é o valor crítico da somatória (WALLER; GOTWAY, 2004). Silverman (1986) diz que a estimação de *kernel* ( $\tilde{f}$ ) pode ser considerada uma soma de nós, cada nó representa um evento ou caso observado. A função de *kernel* vai determinar o formato dos nós e o parâmetro de suavização ( $b$ ) determina a largura dos nós como mostra a Figura 5.

**Figura 5** - Estimativa de *Kernel*.

**Fonte:** Adaptado de (SILVERMAN, 1986)

**Nota:** A largura de banda ( $b$ ) possui raio de 200 metros,  $u$  representa cada caso e a linha azul é a somatória dos *Kernels*.

A densidade de *kernel* foi estimada, aplicando o plugin de heatmap para Qgis, sobre a camada de pontos de casos de hanseníase, para gerar a camada de raster, onde cada pixel da camada possui os valores da função. Nesse trabalho, foi aplicada a função quártica de *kernel*, que é função configurada com padrão no plugin de heatmap definida por

$$\frac{15}{16b} [1 - (u^*)^2]^2 I(|u^*| \leq 1), \quad (3)$$

onde  $I$  é a função indicadora que atribui o valor 1 se a expressão é verdadeira ou 0 se for falsa,  $b$  é o parâmetro de suavização e  $u^* = \frac{u_0 - u_i}{b_u}$  (BAILEY; GATRELL, 1995; WALLER; GOTWAY, 2004).

A interpolação é ponderada por uma largura de banda, que nesse caso foi definida como um raio de influência de 200 metros de cada ponto, que foi escolhida de forma exploratória, com resolução de 2000 por 2600. O valor de banda deve ser escolhido de modo que detalhes importantes dos dados não sejam perdidos com a suavização exagerada ou com ruído causado pelo valor baixo na banda (SHEATHER; JONES, 1991; SILVERMAN, 1984, 1986).

### 3.3 Simulações de Monte Carlo

A simulação de Monte Carlo proposta por Dwass (1957) é usada para gerar réplicas dos dados de forma aleatória, a fim de obter aproximações de funções complexas contornando a necessidade de uma aproximação de forma analítica.. A estatística é calculada para cada simulação, acumulando as distribuições usadas na comparação com a hipótese para o cálculo do valor-p. A hipótese nula segue o princípio de que, se a distribuição for considerada Poisson, acontece independente do evento anterior, então a média e a variância devem ser aproximadamente iguais e, em qualquer problema prático, o cálculo de ambas servirá como um guia para a adequação das hipóteses. A comparação da hipótese é feita para saber se o valor-p é excepcionalmente maior ou menor que a distribuição simulada, dada por,

$$P_{maior} = \frac{N_{maior} + 1}{N_{sim} + 1} \quad P_{menor} = \frac{N_{menor} + 1}{N_{sim} + 1} \quad (4)$$

onde o  $N_{sim}$  é o número de simulações de Monte Carlo,  $N_{maior}$  são os valores das estatísticas são maiores ou iguais do que os valores observados na hipótese e  $N_{menor}$  são os valores menores ou iguais aos observados na estatística, o valor unitário somado ao número de simulações e ao número dos resultados da estatística das simulações é referente a hipótese observada, por isso, o número de repetições de Monte Carlo é geralmente, 99, 999, 9999... simulações (DURBECK et al., 2012; TURNBULL et al., 1990).

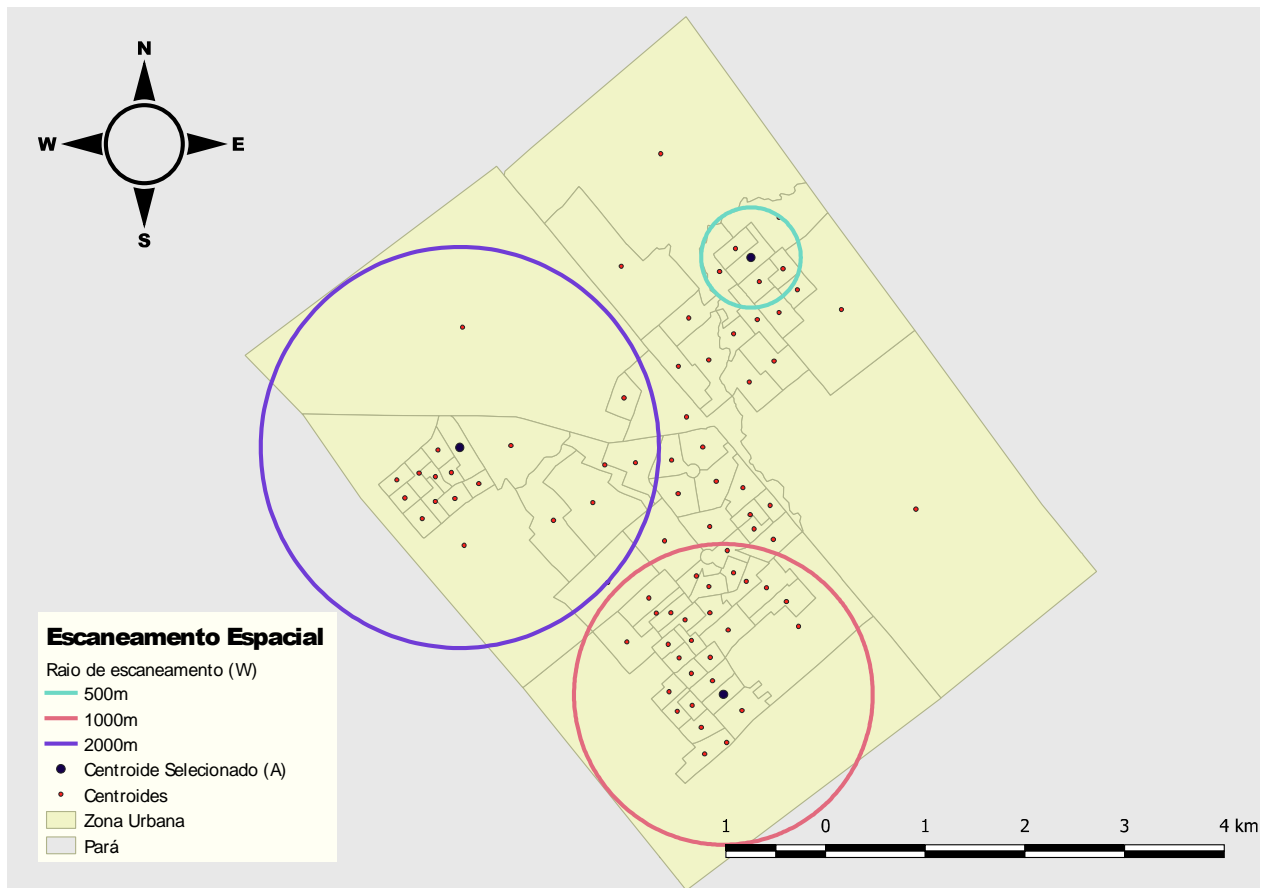
### 3.4 Método de detecção de *cluster* de Kulldorff

O método de Kulldorff é usado para identificar *clusters* em uma distribuição aleatória de casos distribuídos nos setores de determinado espaço geográfico. A detecção é realizada na camada de polígonos, então contorna a necessidade de uma camada de pontos geolocalizados dos casos. Kulldorff (1997) descreve os modelos de escaneamento espacial de Poisson e Bernoulli e um teste de razão de verossimilhança com a hipótese nula para cada modelo. O modelo de Bernoulli é usado em problemas de contagem binária, enquanto o modelo de Poisson é usado para problemas com variáveis

contínuas, e tem a implementação bem mais complexa do que o de Bernoulli. Os dois modelos são similares e o de Poisson pode ser aplicado no mesmo cenário que o Bernoulli com resultados bem próximos. Com isso, o modelo geralmente implementado nos *softwares* é o de Poisson.

O teste faz uma varredura em formato de zonas circulares  $W$  iniciando no centroide de uma subdivisão ( $A$ ) dentro da zona de estudo, representada na Figura 6, considerando que essa região é particionada em subdivisões geográficas, em cada subdivisão, temos os casos da doença e a população. As zonas circulares vão variando de forma contínua ou em passos definidos pelo pesquisador, o raio inicial da zona é definido e varia seu tamanho até que alcance tamanho máximo definido ou que consiga englobar um determinado percentual, que não pode ser maior do que 50%, da população (KULLDORFF; NAGARWALLA, 1995).

**Figura 6** - Representação gráfica da varredura espacial.



**Fonte:** Adaptado de (DURBECK et al., 2012)

**Nota:** No algoritmo representado acima, os círculos variam do menor para o maior círculo em todos os centroides, um centroide por vez. No caso acima, foram 4 passos começando de 500m até 2000m.

Naturalmente fronteiras das subdivisões geográficas são irregulares, então as zonas agregam indivíduos da mesma subdivisão que estejam fora da zona circular, da mesma forma que excluem indivíduos de outra subdivisão que estão dentro da zona circular, isso é feito para verificar se a divisão estudada faz parte de um *cluster* sem levar em conta o formato das subdivisões e comparando com os dados que estão fora da janela de escaneamento. Como o tamanho das zonas circulares são variáveis, o teste estatístico não poderá ser feito usando o número máximo de pontos da zona em determinado tempo, então usa-se o teste estatístico de razão de verossimilhança (KULLDORFF, 1997; LOADER, 1991). Enquanto o raio da zona varia, o teste de significância usando simulações de Monte Carlo é feito para saber se aquela zona corresponde a um *cluster* para a distribuição no modelo de Poisson temos,

$$\frac{L(W)}{L_0} = \frac{\left(\frac{X(W)}{\mu(W)}\right)^{X(W)} \left(\frac{X(A\setminus W)}{\mu(A\setminus W)}\right)^{X(A\setminus W)}}{\left(\frac{X(A)}{\mu(A)}\right)^{X(A)}} \quad (5)$$

se  $X(W)/\mu(W) > X(A\setminus W)/\mu(A\setminus W)$ , e  $L(W) = 1$  caso contrário, onde  $X$  é o número aleatório de eventos observados,  $A$  área onde o caso ocorre (pode ser um vetor espacial ou temporal),  $L(W)$  é a função de verossimilhança para  $W$ ,  $W$  é a coleção de zonas (janela) em  $A$ ,  $L_0$  é a função de verossimilhança para a hipótese nula e  $\mu$  número de eventos esperados.

O teste se dá com o círculo do escaneamento faz uma coleção de zonas  $W$ , onde  $\mu$  define o número de casos esperados em  $A$ , representando uma taxa que gera eventos sob a hipótese nula.  $L(W)$  é a função de verossimilhança para a zona  $W$  e  $L_0$  é a função de verossimilhança para a hipótese nula. A expressão acima mostra que o número de eventos dentro do círculo é maior do que o esperado quando comparado com o que está fora do círculo (KULLDORFF, 1997, 1999).

O *software* escolhido para análises foi o ClusterSeer 2.5, que implementa o modelo de Poisson pois ele funciona melhor em situações de contagem de casos e população em risco. A configuração escolhida para estabelecer a vizinhança de cada setor foi a de contiguidade *Queen*, que seleciona como vizinho todos os setores que tocam as fronteiras do setor observado. E a janela de escaneamento é definida pelo usuário que informa o raio inicial e final do círculo assim como os passos da variação. Após o

processamento dos dados, o *software* retorna o mapa com os três *clusters* mais prováveis e seus respectivos valores-p de verossimilhança (DURBECK et al., 2012).

### 3.5 Indicador de autocorrelação espacial

A autocorrelação espacial é usada para indicar a existência de *clusters* onde os valores das taxas observadas dentro de uma divisão espacial, são similares ou não em relação aos seus vizinhos. O Indicador de autocorrelação espacial é classificado com o termo “global” quando se refere ao uso de dados de todo o espaço geográfico analisado; e “local” quando se refere a população e casos observados em uma subdivisão, setor ou distrito do espaço geográfico analisado.

A equação básica de autocorrelação espacial (equação (6)), calcula a similaridade (*sim*) dos valores (*Y*) nas localizações *i* e *j*, então pondera (peso *p*) dessa similaridade pela proximidade das localizações *i* e *j* (WALLER; GOTWAY, 2004) dado por

$$\frac{\sum_{i=1}^N \sum_{j=1}^N p_{ij} sim_{ij}}{\sum_{i=1}^N \sum_{j=1}^N p_{ij}}, \quad (6)$$

onde *N* é o total de localizações, *sim* é a similaridade dos valores (*Y*), *i* e *j* são as coordenadas das localizações e *p* é o peso (distância)

#### 3.5.1 Global

O trabalho de Moran (1948, 1950) foi usado como base para aplicação em estatística espacial, conhecido como I de Moran, que compara as similaridades de atributos, localização e distribuição em mapas estatísticos, que informa o nível de interdependência entre todos os espaços geográficos estudados, a fim de determinar se no espaço estudado há uma distribuição aleatória de dados, ou se os dados estão distribuídos de uma forma especial.

O I de Moran segue a equação básica com a similaridade entre as regiões *i* e *j* definidas pelo produto da diferença entre *Y<sub>i</sub>* e *Y<sub>j</sub>* com a média geral,

$$sim_{ij} = (Y_i - \bar{Y})(Y_j - \bar{Y}) \quad (7)$$

De acordo com Anselin (1993) apesar dos vários métodos de análises geográficas, somente algumas são apropriadas para lidar explicitamente com uma larga escala de dados. A melhor forma de calcular o I de Moran para grandes quantidades de dados é computacionalmente, onde o *software* computa o I de Moran com o objetivo de testar a hipótese de interdependência comparando com uma hipótese nula, a hipótese nula é feita com uma randomização dos dados espaciais repetida várias vezes, um valor-p é gerado, também calcula-se a variância do mapa real, então a média dos valores I de Moran devem ser divididos pela raiz quadrada da variância a fim de obter o valor-z . Quando o índice de Moran é positivo e é estatisticamente significativo, temos formação de *cluster*, no caso do I de Moran negativo, também estatisticamente significativo, não há formação de *clusters*, e em qualquer caso se o teste não for estatisticamente significativo, temos uma distribuição espacial totalmente aleatória.

O *software* escolhido para gerar o correlograma do índice de Moran global foi o *Point Pattern Analysis* (PPA), as variáveis testadas foram o número de casos e a taxa de incidência em cada setor censitário. A distribuição espacial é feita usando as coordenadas dos centroides de cada setor e o teste de vizinhança é feito usando um raio que começa com uma distância de 500 metros e faz 20 incrementos até chegar em 10 mil metros.

### 3.5.2 Local

O indicador local de autocorrelação espacial, denominado LISA (*local Indicator of Spatial Association*) por Anselin (1995), deriva do método global com a seguinte forma básica para cada região *I*:

$$\sum_{j=1}^N p_{ij} sim_{ij} \quad (8)$$

O método LISA usa espaços geográficos em forma de polígonos, e os vizinhos de cada divisão espacial pode ser calculado usando pesos espaciais ou matrizes de contiguidade (WALLER; GOTWAY, 2004). Anselin (1995) sugere que o indicador local de autocorrelação espacial é qualquer estatística que mostra *clusters* espaciais de valores similares em torno da observação; E a soma dos LISAs de todas as observações

é proporcional ao indicador global de autocorrelação espacial. LISA é usado para descobrir onde estão localizados os *clusters* e hotspots indicados pelo método global. Assim como no método global, a melhor forma de fazer a inferência é computacionalmente, o método é o mesmo do global, só que no LISA ele é feito em cada divisão espacial estudada, então descobrimos quais localizações são significantes estatisticamente, assim como no global, se o local estudado não for significativo, os dados estão distribuídos de forma aleatória e não há o que concluir com o LISA.

A versão local do método I de Moran é, provavelmente, o mais usado dos LISA, é aplicado na *i*-ésima região como

$$I_i = (Y_i - \bar{Y}) \sum_{j=1}^N p_{ij} (Y_j - \bar{Y}) \quad (9)$$

E no trabalho de Anselin (1993) sugere o uso de gráficos de dispersão para visualizar a autocorrelação espacial I de Moran. O gráfico é centrado em zero no eixo x e y e tem quatro quadrantes que representam diferentes tipos de associação entre os valores de determinada localização e suas variações espaciais, que são a média ponderada dos valores nas localizações vizinhas. No quadrante superior temos valores altos (alto-alto), e no quadrante inferior valores baixos (baixo-baixo). O canto superior direito e o inferior esquerdo representam uma associação espacial positiva, ou seja, determinado lugar tem vizinhos com valores parecidos. O quadrante superior esquerdo (baixo-alto) são espaços com valores altos cercados de valores baixos e inferior direito (alto-baixo) são valores baixos cercados de valores altos, logo esses quadrantes correspondem a associação espacial negativa.

O Geoda foi o *software* escolhido para o mapeamento dos *clusters* e visualização do gráfico de dispersão do método I de Moran Local, o além do mapa de *clusters* é gerado um mapa de significância, com a combinação dos dois mapas é possível descobrir quais *clusters* são estatisticamente significantes (ANSELIN; SYABRI; KHO, 2006).

### 3.6 Método de detecção de interações espaço-temporais de Knox

De acordo com Knox (1964), se uma epidemia é representada em um gráfico em relação ao tempo, os eventos podem aparecer de forma muito dispersa, e algumas

unidades do tempo podem ficar vazias, então o conceito de epidemia poderá perder o sentido em relação a sua definição e ser desconsiderado, pois distribuição geográfica é um dos elementos que devem ser considerados.

O método de Knox (1964) usa o número de pares de casos  $X$  que estão próximos em tempo e distancia estabelecidos, para identificação de *cluster* espaço-temporal, definidos por

$$X = \sum_{i=1}^N \sum_{j=1}^{i-1} s_{ij} t_{ij} \quad (10)$$

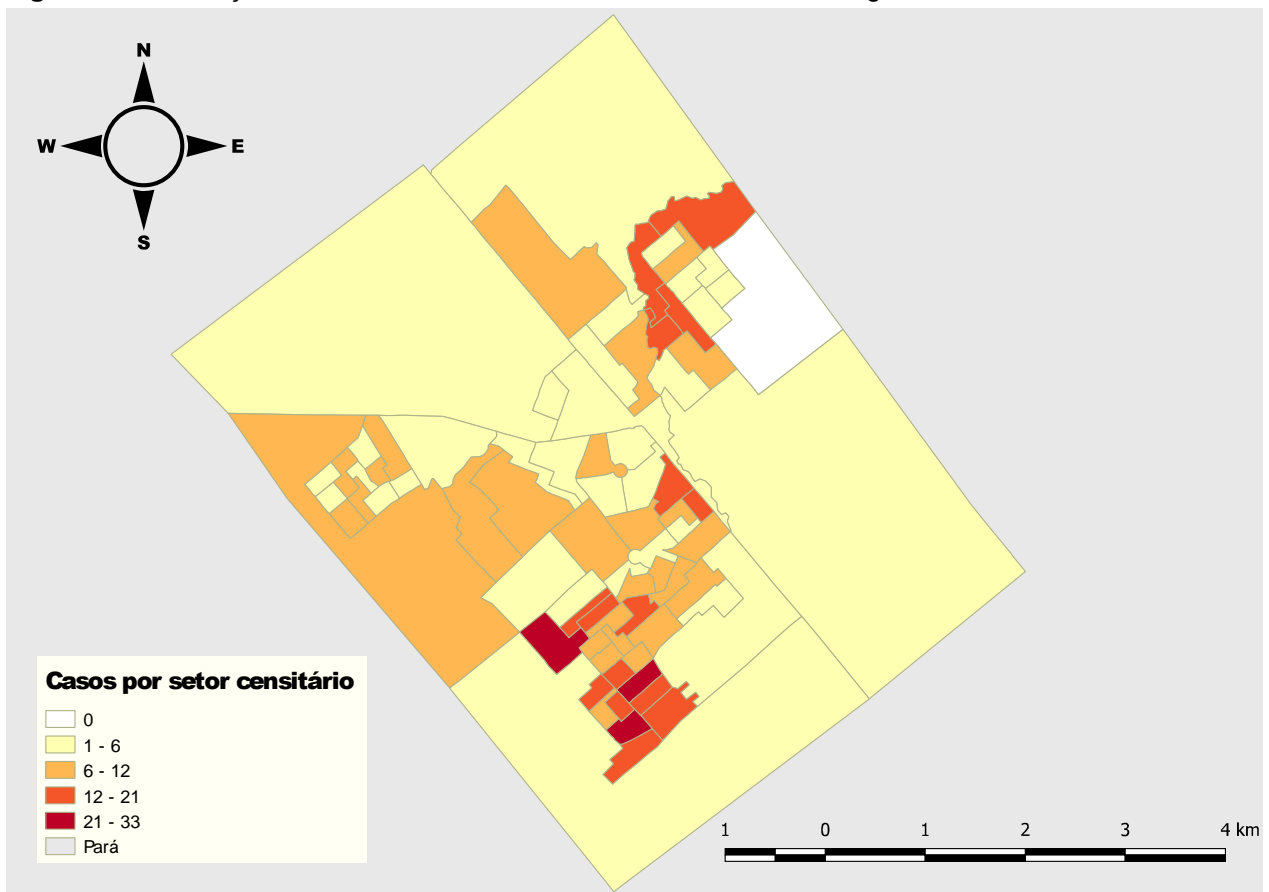
onde  $N$  é o número total de casos;  $s_{ij}$  indica com 1 se a distância espacial entre os pares de caso estão dentro do limite estabelecido e 0 se não estiver; e  $t_{ij}$  indica com 1 se os casos ocorrem dentro do limite de tempo estabelecido e 0 caso contrário.

O *software* utilizado foi o *ClusterSeer2*, que testa a distribuição espacial e temporal dos casos  $X$ , usando o teste  $\chi^2$  (qui-quadrado) para calcular a probabilidade da classificação dos dados em perto no tempo e perto na distância; longe no tempo e perto na distância; perto no tempo e longe na distância e longe no tempo e longe na distância. As simulações de Monte Carlo, são calculadas dentro de uma hipótese nula, gerando um valor-p de significância baseado na comparação dos valores observados com os valores esperados em  $X$  (DURBECK et al., 2012).

## 4 RESULTADOS

De acordo com a base de dados do SINAN, 1047 casos novos foram notificados no período estudado em Paragominas, 791 só na zona urbana e 720 desses casos foram mapeados. Foi observada uma média de nove casos por setor censitário, com no mínimo zero casos e máximo de 33 casos em cada setor e a Figura 7 mostra a distribuição da hanseníase na zona urbana de Paragominas.

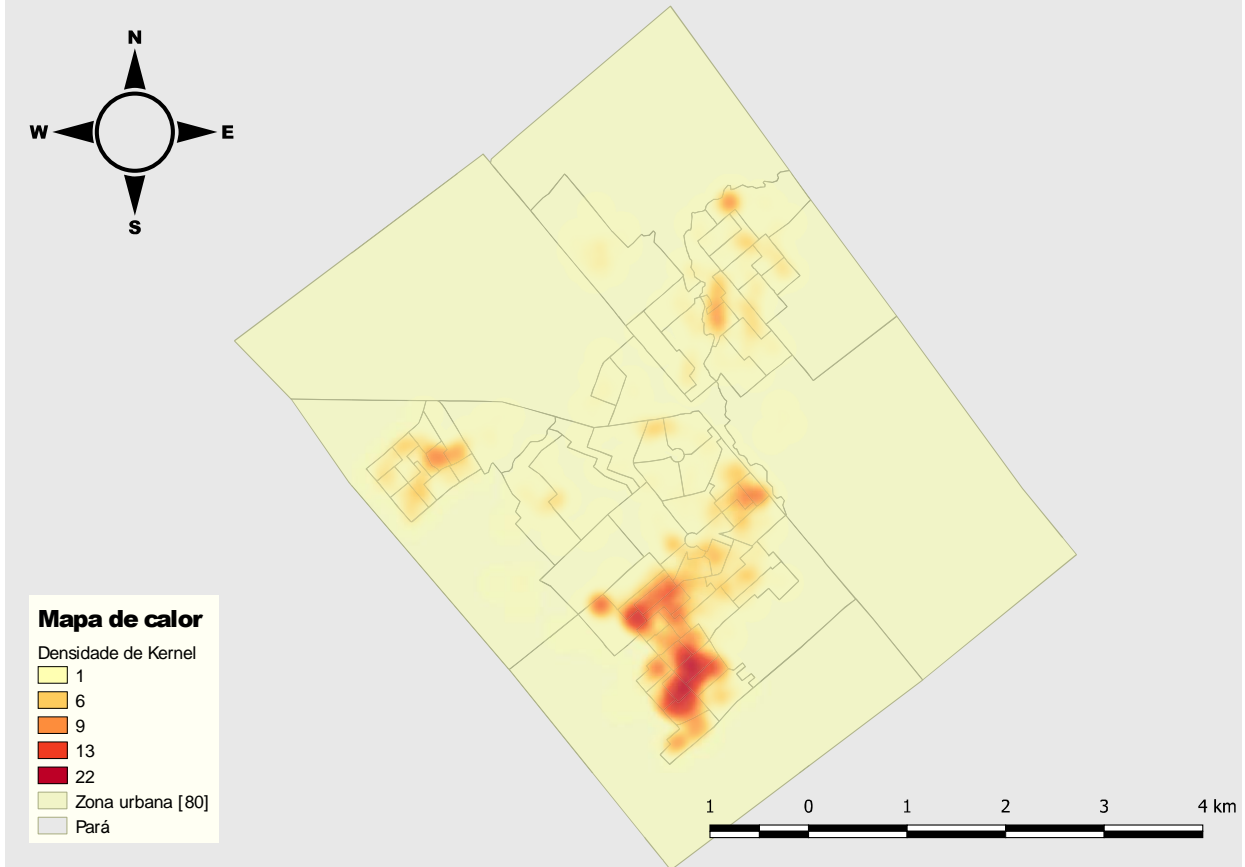
**Figura 7** - Distribuição de casos de Hanseníase na zona urbana de Paragominas, Pará, 2004-2010.



Fonte: Autor (2018)

A Figura 8 mostra o mapa de calor e podemos observar a densidade de casos num raio de 200 metros e também a concentração de casos em determinadas áreas da cidade e a heterogeneidade na distribuição de casos pelo município.

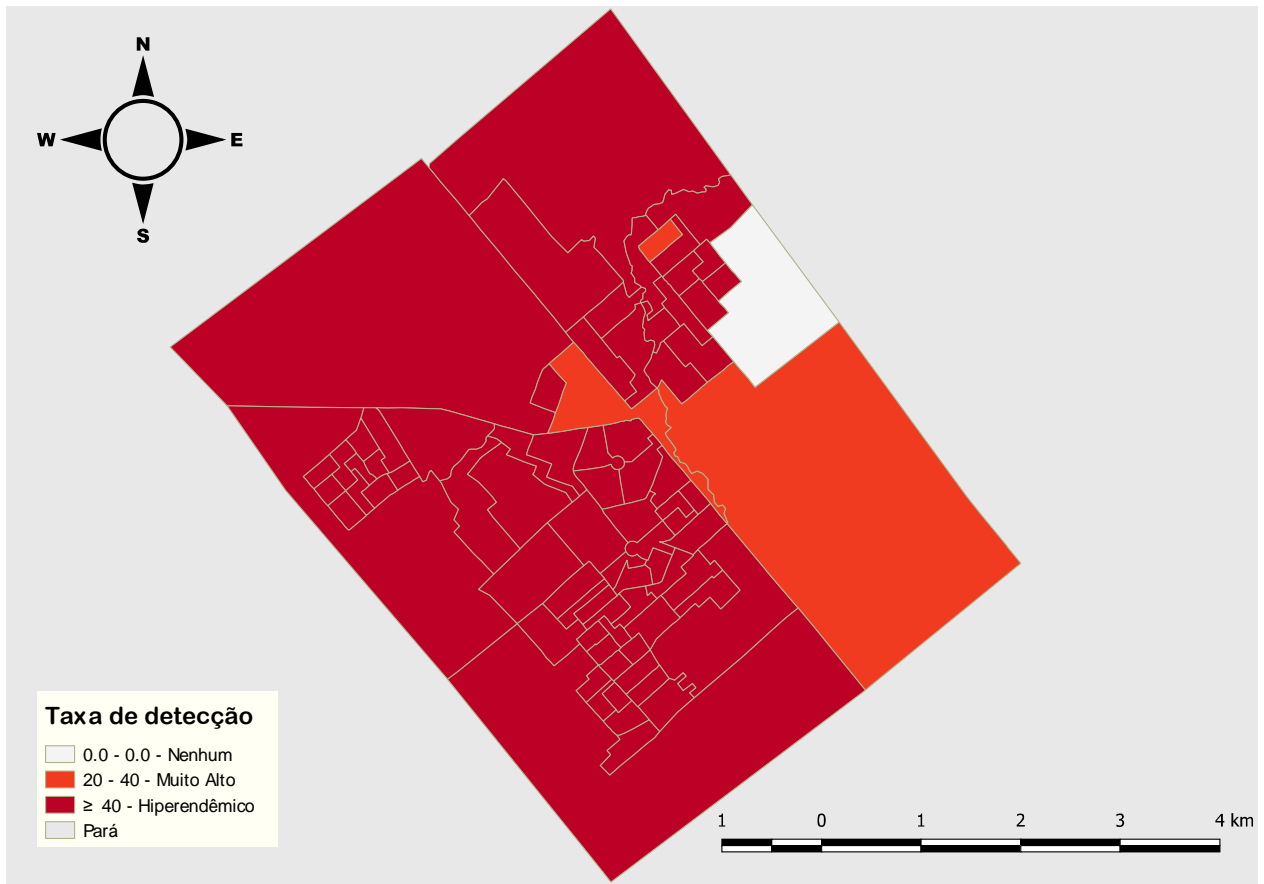
**Figura 8** - Mapa de calor dos casos de hanseníase na zona urbana de Paragominas, Pará, 2004-2010.



**Fonte:** Autor (2018)

A Figura 9 mostra que 76 dos 80 setores censitários urbanos foram classificados como hiperendêmicos para hanseníase. Para o cálculo da taxa em Paragominas, foi considerado a média de casos no período de 2004 a 2010, pela população de cada setor censitário,

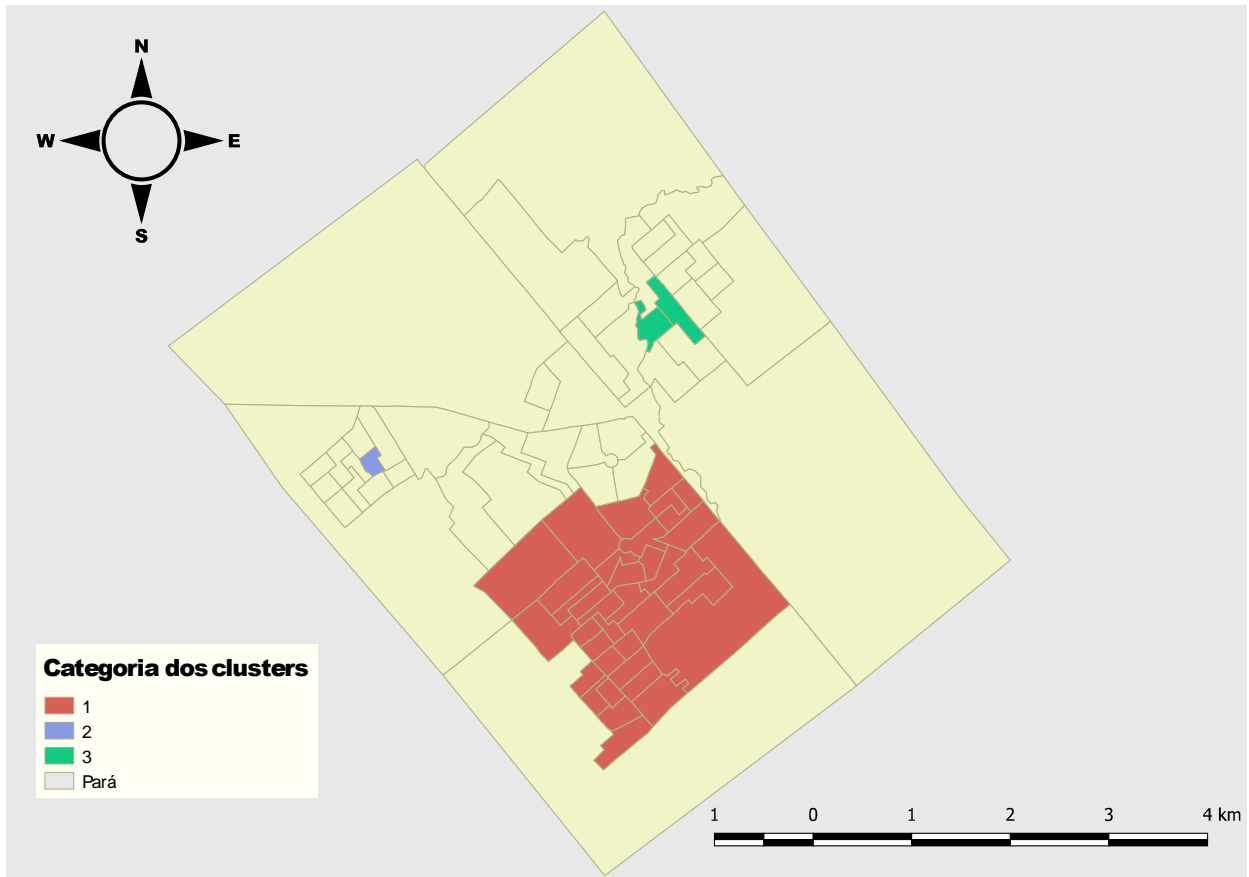
**Figura 9** - Taxa de detecção de casos novos de hanseníase, na zona urbana de Paragominas, Pará, 2004-2010.



**Fonte:** Autor (2018)

O mapa da Figura 10 foi produzido com a varredura espacial de Kulldorff com o objetivo de determinar formações de *Clusters* na zona urbana de Paragominas com subdivisões geográficas definidas por setores censitários.

**Figura 10** - Método de Kulldorff para detecção de *clusters* de casos de hanseníase, na zona urbana de Paragominas, Pará, 2004-2010.



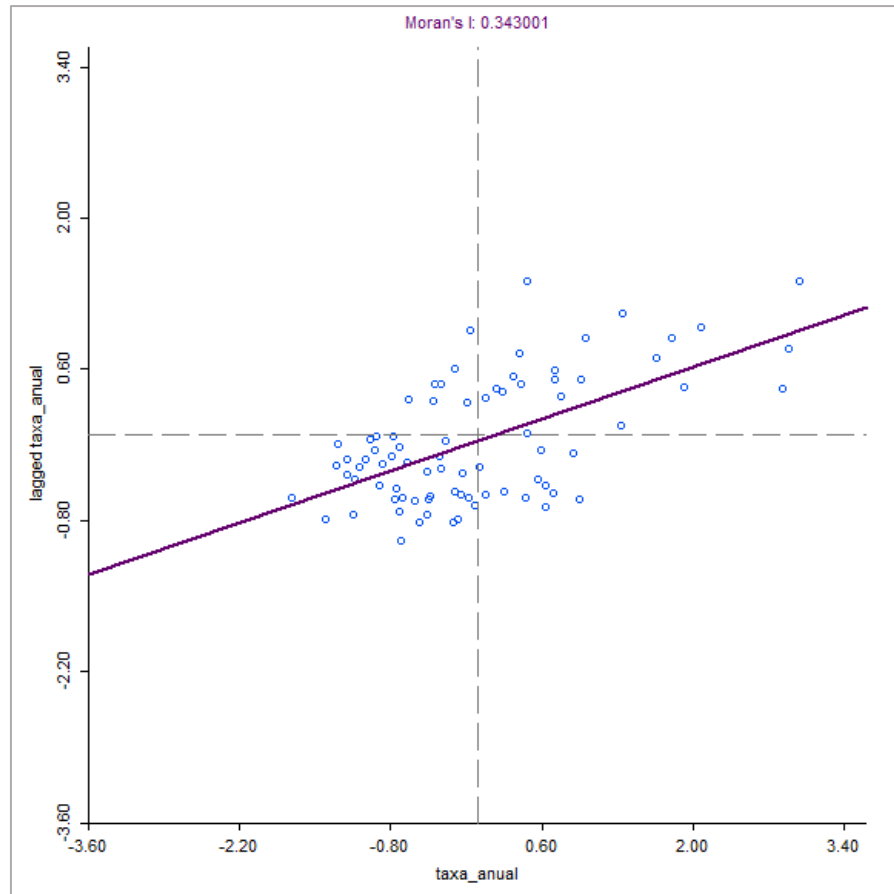
Fonte: Autor (2018)

Para investigar o agrupamento espacial, foi utilizada a autocorrelação espacial global de Moran na taxa de detecção dos casos de hanseníase, na Tabela 1 os valores observados são comparados com os valores esperados dentro da simulação de Monte Carlo (999 interações).

Tabela 1 - Autocorrelação espacial global de Moran para a taxa de detecção anual.

| <b>Distancia (metros)</b> | <b>I de Moran</b> | <b>I esperado</b> | <b>Variância</b> | <b>Valor Z</b> |
|---------------------------|-------------------|-------------------|------------------|----------------|
| 500                       | 0,4222            | -0,0127           | 0,0051           | 6,0903         |
| 1000                      | 0,3527            | -0,0127           | 0,0018           | 8,7134         |
| 1500                      | 0,2837            | -0,0127           | 0,0011           | 9,1248         |
| 2000                      | 0,2586            | -0,0127           | 0,0008           | 9,8092         |
| 2500                      | 0,2302            | -0,0127           | 0,0006           | 9,8948         |
| 3000                      | 0,1997            | -0,0127           | 0,0005           | 9,6846         |
| 3500                      | 0,1796            | -0,0127           | 0,0004           | 9,6684         |
| 4000                      | 0,1647            | -0,0127           | 0,0004           | 9,4488         |
| 4500                      | 0,1586            | -0,0127           | 0,0003           | 9,378          |
| 5000                      | 0,1562            | -0,0127           | 0,0003           | 9,3305         |
| 5500                      | 0,1553            | -0,0127           | 0,0003           | 9,2959         |
| 6000                      | 0,1549            | -0,0127           | 0,0003           | 9,2786         |
| 6500                      | 0,1548            | -0,0127           | 0,0003           | 9,2726         |
| 7000                      | 0,1548            | -0,0127           | 0,0003           | 9,2726         |
| 7500                      | 0,1548            | -0,0127           | 0,0003           | 9,2726         |
| 8000                      | 0,1548            | -0,0127           | 0,0003           | 9,2726         |
| 8500                      | 0,1548            | -0,0127           | 0,0003           | 9,2726         |
| 9000                      | 0,1548            | -0,0127           | 0,0003           | 9,2726         |
| 9500                      | 0,1548            | -0,0127           | 0,0003           | 9,2726         |
| 10000                     | 0,1548            | -0,0127           | 0,0003           | 9,2726         |

A análise local de associação espacial I de Moran mostra quais setores censitários possuem semelhança entre si, podemos ver no gráfico da Figura 11, cada ponto corresponde a um setor, e no quadrante direito superior temos os setores que são similares para a formação de um *cluster*.

**Figura 11** - Gráfico de dispersão de I de Moran

**Fonte:** Autor (2018)

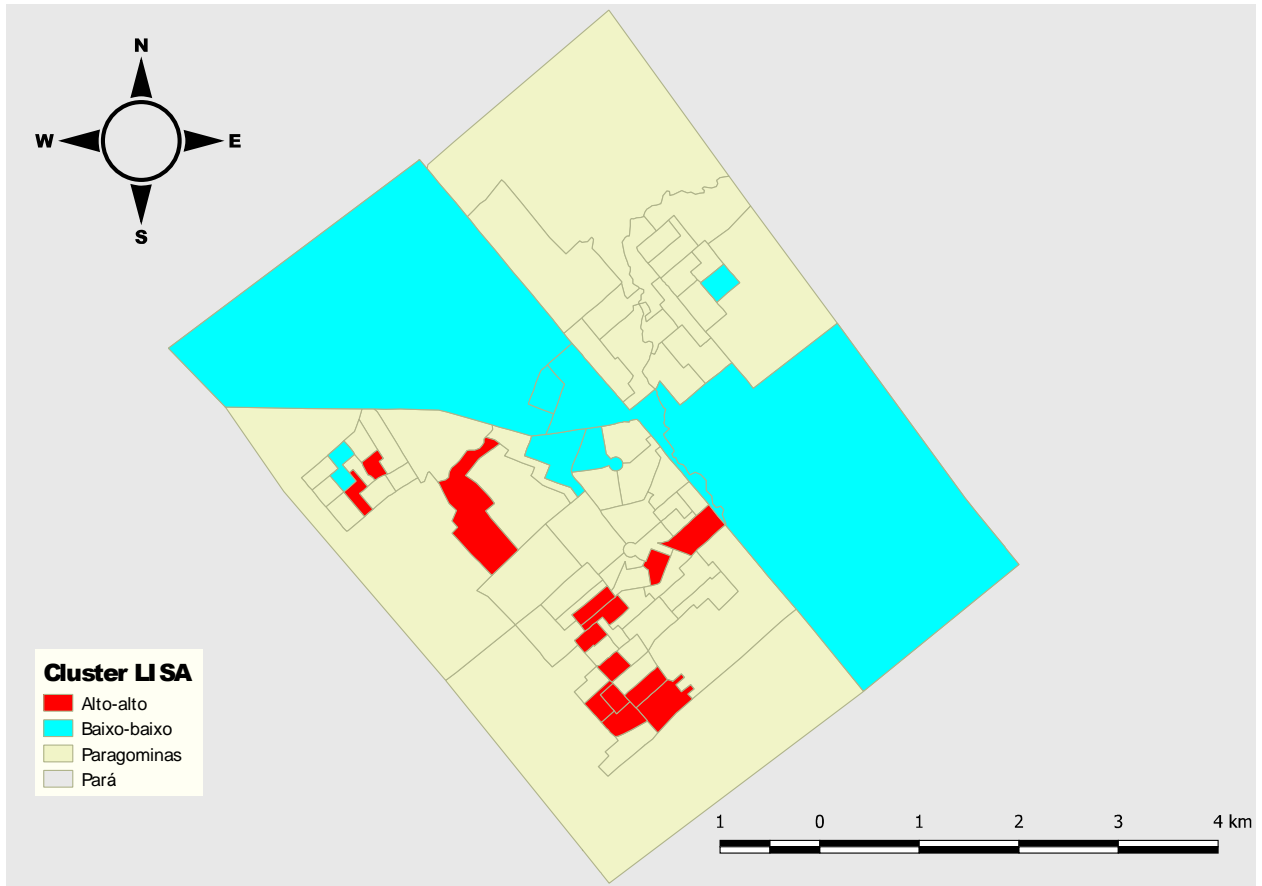
Nem todos os setores são estatisticamente significantes, o mapa da Figura 12 mostra quais setores são estatisticamente significantes para a análise local de associação espacial.

**Figura 12** - Mapa de significância LISA

**Fonte:** Autor (2018)

Com a combinação do gráfico de dispersão e o mapa de significância, o mapa da Figura 13 mostra *clusters* estatisticamente significante, entre valores semelhantes espacialmente, isto é, entre setores com altas taxas de detecção (alto-alto) e entre os setores com baixas taxas de detecção (baixo-baixo).

**Figura 13** - *Cluster LISA* de casos de hanseníase, na zona urbana de Paragominas, Pará, 2004-2010.



**Fonte:** Autor (2018)

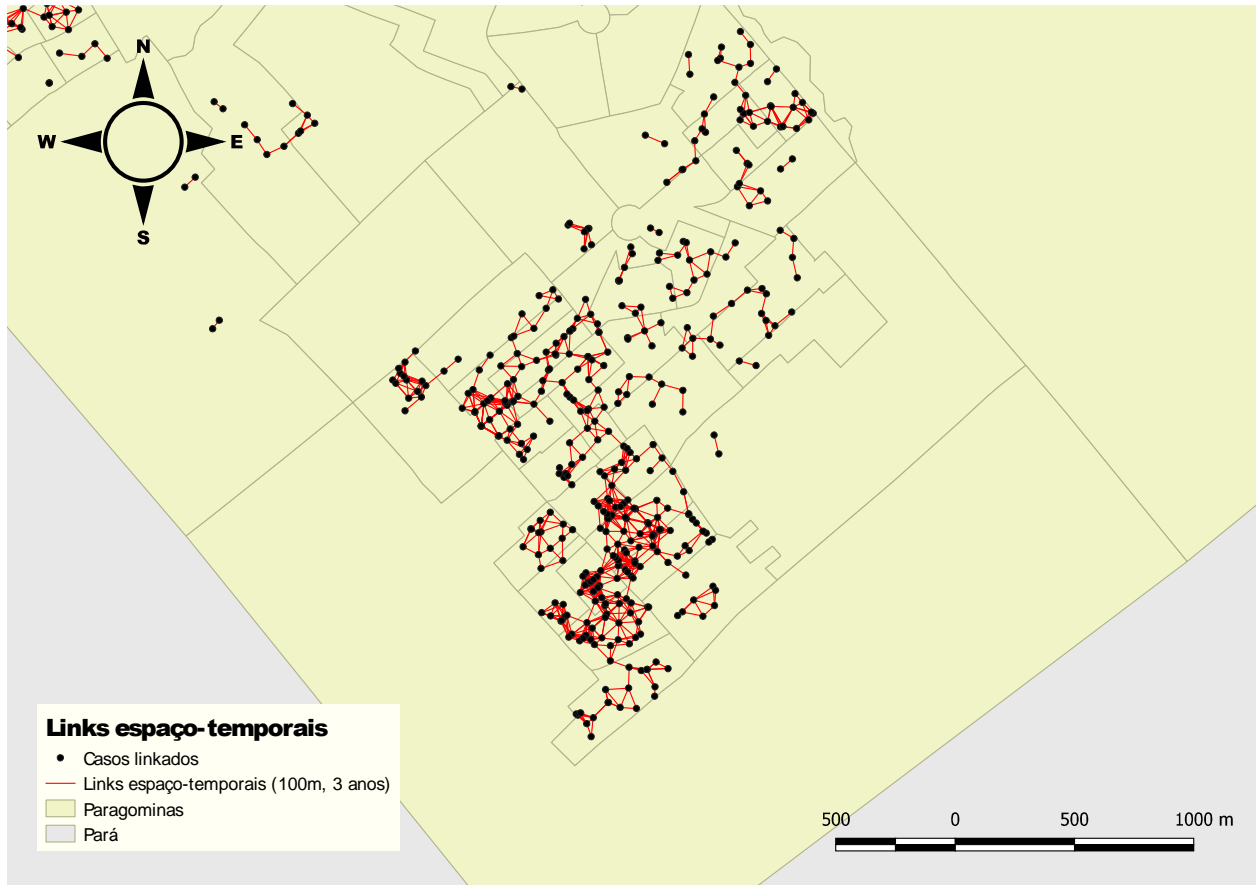
A Tabela 2 exibe os resultados da análise de *Cluster* espaço-tempo Knox para os casos de hanseníase baseados em diferentes faixas de espaço-tempo, com isso podemos identificar se há ligação entre os casos tanto no espaço quanto no tempo.

Tabela 2 - Análise espaço temporal de Knox para casos de hanseníase.

| <b>Faixa espaço-temporal (Metros-Anos)</b> | <b>Número de Links espaço-temporais</b> | <b>Número de casos</b> | <b>Valor de <math>p</math> (999 simulações de Monte Carlo)</b> |
|--|---|------------------------|--|
| 50 - 1                                     | 226                                     | 273                    | 0,002  |
| 50 - 2                                     | 336                                     | 319                    | 0,003  |
| 50 - 3                                     | 359                                     | 375                    | 0,087  |
| 100 - 1                                    | 743                                     | 551                    | 0,05   |
| 100 - 2                                    | 1134                                    | 573                    | 0,079  |
| 100 - 3                                    | 1393                                    | 605                    | 0,278  |
| 200 - 1                                    | 2577                                    | 664                    | 0,006  |
| 200 - 2                                    | 3909                                    | 683                    | 0,041  |
| 200 - 3                                    | 4792                                    | 694                    | 0,348  |

A análise que faz o link de 100 metros no período de 3 anos revela que 83% dos casos tem conexão com algum outro caso, a Figura 14 mostra em detalhe os casos de hanseníase dentro da área do *cluster* principal.

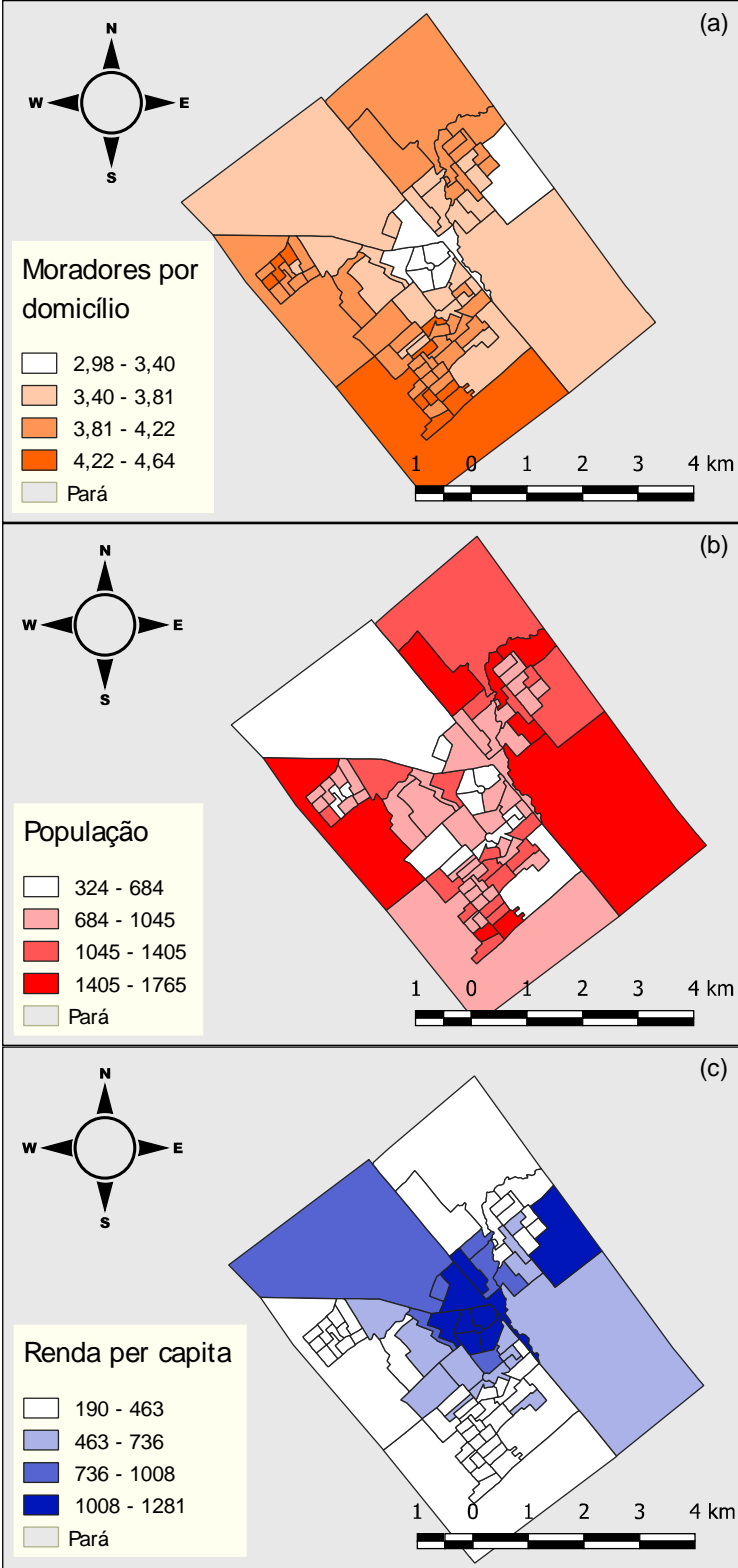
**Figura 14** - Método de Knox para detecção de links espaço-temporais, no *cluster* principal da zona urbana de Paragominas, Pará, 2004-2010.



**Fonte:** Autor (2018)

A partir dos dados obtidos pelo senso de 2010 do IGBE foi elaborado o mapa da distribuição de renda per capita, na Figura 15, com o objetivo de ligar os casos com a baixa renda da população, e mapas da densidade de pessoas por casas e população total em cada setor.

**Figura 15** - Distribuição de renda, população e taxa de moradores por residência em Paragominas, Pará, 2010.



Fonte: Autor (2018)

## 5 DISCUSSÃO

O uso dos SIGs mostrou-se uma eficiente ferramenta para o entendimento da epidemiologia da hanseníase no município de Paragominas, seu uso é estimulado pela OMS e sem a ajuda dos SIGs e do GPS o mapeamento seria mais demorado, complicado e algumas estatísticas não poderiam ser realizadas. A revisão bibliográfica feita para entender as fórmulas e métodos estatísticos usados nos SIGs ajuda na interpretação dos resultados obtidos, assim como evita equívocos na conclusão dos resultados estatísticos e como eles representam a distribuição da doença no município. A distribuição espacial da hanseníase é heterogênea com presença de *clusters* ao sul da zona urbana de Paragominas.

O período estudado limita o entendimento da doença para os dias atuais, porém o resultado das análises são uma importante ferramenta para elaboração de ações de controle da doença focadas em setores específicos de Paragominas. Nesse trabalho as estatísticas foram realizadas somente com base nos dados brutos, em trabalhos como o de Barreto et al. (2014) as análises foram feitas também nos dados suavizados, e como alguns autores apontam, os dados suavizados podem ajudar a estimar valores que podem estar perdidos em certas áreas além de suavizar variáveis aleatórias (MARTIN et al., 2003).

Alguns trabalhos (MOURA et al., 2008; OSKAM; SLIM; BÜHRER-SÉKULA, 2003; SPENCER et al., 2012) demonstram que a sorologia do anti-glicolípido fenólico (PGL-I) pode ser usada para classificação de pacientes, identificação de possíveis recidivas, monitoramento dos casos e identificação do risco dos contatos intradomiciliares desenvolverem a doença. Nos trabalhos de Barreto et al. (2014, 2015) a sorologia do PGL-I é usada para melhorar a detecção precoce de casos de hanseníase tentar identificar possíveis focos de transmissão usando SIG para análise espacial em combinação com a sorologia.

Alguns autores (BARRETO et al., 2014; LAPA et al., 2006; QUEIROZ et al., 2010) aplicam vários métodos estatísticos para contornar limitações e desvantagens e unir os resultados para uma visão privilegiada da distribuição da hanseníase que não é possível ser observada usando somente um método. As análises espaciais dos casos de

hanseníase, no período de 2004 a 2010, mostraram que a doença se comporta de forma heterogênea no município, 95% dos setores censitários são hiperendêmicos. A análise de *kernel* e a contagem de casos por setor mostram uma concentração de casos ao sul da zona urbana de Paragominas, entretanto a essa estimativa não provê nenhuma significância estatística das concentrações observadas.

Com o método de kulldorff, para escaneamento espacial, identificamos a formação de clusters significantes estatisticamente. No trabalho de Hjalmar et al., (1996) O método de kulldorff é aplicado para detecção de *cluster* de leucemia entre crianças na Suécia, e para o entendimento dos resultados os autores mostram as fórmulas, algoritmos e programas usados durante a varredura de detecção de *cluster*. Nenhum cluster significativo foi encontrado.

Os índices de autocorrelação espacial global e local mostraram que alguns setores de Paragominas possuem similaridades entre si, em relação a taxa de detecção da hanseníase, e a análise LISA com o método I de Moran mostrou clusters significantes estatisticamente de setores semelhantes no município. Outro exemplo da aplicação do índice de autocorrelação no trabalho de Flahaut et al., (2003) que tenta identificar e delimitar locais na estrada com concentração de acidentes, comparando os números dos índices a densidade de kernel para tentar determinar as chances de um acidente ocorrer em cada localização identificada.

A análise de clusters espaço-temporais de Knox mostrou resultados um tanto quanto inesperados, tivemos números muito altos de ligações espaço-temporais entre os casos, com a menor janela ( espaço de 50 metros dentro de um ano) tivemos 37,91% dos casos ligados significantes estatisticamente, porém com a janela de 100 metros e 3 anos teve baixa significância estatística acreditamos que isso ocorreu devido ao alto número de ligações entre os casos (83%). A tabela de janelas da análise de Knox pode ajudar a entender como a doença foi transmitida entre alguns moradores de Paragominas. Gilman e Knox (1995) identificaram *clusters* significativos de linfoma e leucemia entre crianças, principalmente as nascidas com 1km e 5 meses de distância, as diferentes *clusters* mostram que os casos podem ter sido expostos a mais de um tipo de agente causador das doenças.

Esse trabalho pode ser considerado introdutório para estudantes de SIG ou de estatística pois as técnicas e métodos descritos podem ser aplicados para o estudo e planejamento de ações para hanseníase e outras doenças endêmicas ou dados relacionados a saúde em diversos municípios, estados ou países.

## 6 CONCLUSÃO

As análises espaciais da hanseníase mostraram que a doença é distribuída de forma heterogênea na zona urbana do município de Paragominas. Os mapas gerados junto com a utilização de ferramentas para a análise estatística, mostrou a formação de *clusters* nas áreas mais pobres da cidade. O alto número de ligações espaço-temporais observadas em Paragominas ajuda a entender a transmissão da doença ao longo do tempo e entre os vizinhos localizados dentro do raio de análise. Os resultados das análises em conjunto com o uso dos sistemas de informação geográfica para visualização dos mapas ajudam a entender como a doença se comporta dentro do município auxiliando na criação de estratégias de detecção de casos ocultos, realização de ações de controle específicas para o município, especialmente nas áreas de maior risco de transmissão com o objetivo de alcançar a meta da Organização Mundial da Saúde de eliminação da hanseníase como problema de saúde pública no Pará.

## REFERÊNCIAS

ALDSTADT, J.; GETIS, A. Point Pattern Analysis in an ArcGIS Environment. **Published by San Diego State University**, 2002.

ANSELIN, L. **The Moran scatterplot as an ESDA tool to assess local instability in spatial association**. [s.l.] Regional Research Institute, West Virginia University Morgantown, WV, 1993.

ANSELIN, L. Local indicators of spatial association—LISA. **Geographical analysis**, v. 27, n. 2, p. 93–115, 1995.

ANSELIN, L. GeoDa 0.9 user's guide. **Spatial Analysis Laboratory (SAL). Department of Agricultural and Consumer Economics, University of Illinois, Urbana-Champaign, IL**, 2003.

ANSELIN, L.; GETIS, A. Spatial statistical analysis and geographic information systems. **The Annals of Regional Science**, v. 26, n. 1, p. 19–33, 1992.

ANSELIN, L.; SYABRI, I.; KHO, Y. GeoDa: an introduction to spatial data analysis. **Geographical analysis**, v. 38, n. 1, p. 5–22, 2006.

BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. [s.l.] Longman Scientific & Technical Essex, 1995. v. 413

BARRETO, J. G. et al. Spatial analysis spotlighting early childhood leprosy transmission in a hyperendemic municipality of the Brazilian Amazon region. **PLoS neglected tropical diseases**, v. 8, n. 2, p. e2665, 2014.

BARRETO, J. G. et al. Spatial epidemiology and serologic cohorts increase the early detection of leprosy. **BMC infectious diseases**, v. 15, n. 1, p. 527, 2015.

BONHAM-CARTER, G. F. Geographic information systems for geoscientists-modeling with GIS. **Computer methods in the geoscientists**, v. 13, p. 398, 1994.

BRASIL. **TabNet Win32 3.0: Acompanhamento dos dados de Hanseníase**. Disponível em: <<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinannet/hansenia/cnv/hanswbr.def>>. Acesso em: 28 mar. 2019.

CREE, I. A.; SMITH, W. C. Leprosy transmission and mucosal immunity: towards eradication? **Lepr Rev**, v. 69, p. 112–121, 1998.

DURBECK, H. et al. ClusterSeer User Manual book 1 version 2. 5. **BioMedware geospatial research and software**, 2012.

DWASS, M. Modified randomization tests for nonparametric hypotheses. **The Annals of Mathematical Statistics**, p. 181–187, 1957.

ELLIOTT, P.; WARTENBERG, D. Spatial epidemiology: Current approaches and future challenges. **Environmental Health Perspectives**, v. 112, n. 9, p. 998–1006, 2004.

FLAHAUT, B. et al. The local spatial autocorrelation and the kernel method for identifying black zones: A comparative approach. **Accident Analysis and Prevention**, v. 35, n. 6, p. 991–1004, 2003.

GILMAN, E. A.; KNOX, E. G. Childhood Britain. n. March 1994, p. 158–163, 1995.  
HANSEN, G. H. A. Undersøgelser angående spedalskhedens årsager. **Norsk Magazin for Laegeviteuskapen**, v. 4, p. 1–88, 1874.

HJALMARS, U. et al. Childhood Leukaemia in Sweden: Using Gis and a Spatial Scan Statistic for Cluster Detection. **Statistics in Medicine**, v. 15, n. 79, p. 707–715, 1996.

KNOX, E. G. The detection of space-time interactions. **Journal of the Royal Statistical Society. Series C (Applied Statistics)**, v. 13, n. 1, p. 25–30, 1964.

KULLDORFF, M. A spatial scan statistic. **Communications in Statistics-Theory and methods**, v. 26, n. 6, p. 1481–1496, 1997.

KULLDORFF, M. Spatial scan statistics: models, calculations, and applications. In: **Scan statistics and applications**. [s.l.] Springer, 1999. p. 303–322.

KULLDORFF, M.; NAGARWALLA, N. Spatial disease clusters: detection and inference. **Statistics in medicine**, v. 14, n. 8, p. 799–810, 1995.

LAPA, T. M. et al. Análise da demanda de casos de hanseníase aos serviços de saúde através do uso de técnicas de análise espacial. **Cadernos de Saúde Pública**, v. 22, n. 12, p. 2575–2583, 2006.

LAWSON, A. B. **Statistical Methods in Spatial Epidemiology**. [s.l.] Wiley, 2013.

LOADER, C. R. Large-deviation approximations to the distribution of scan statistics. **Advances in Applied Probability**, v. 23, n. 4, p. 751–771, 1991.

MARTIN, S. W. et al. Geographical and temporal distribution of human giardiasis in Ontario, Canada. **International Journal of Health Geographics**, v. 2, n. 1, p. 5, 2003.

MORAN, P. A. P. The interpretation of statistical maps. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 10, n. 2, p. 243–251, 1948.

MORAN, P. A. P. Notes on continuous stochastic phenomena. **Biometrika**, v. 37, n. 1/2, p. 17–23, 1950.

MOURA, R. S. DE et al. Leprosy serology using PGL-I: a systematic review. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 41, p. 11–18, 2008.  
OMS. **Estratégia Global para Hanseníase**. [s.l: s.n.].

OSKAM, L.; SLIM, E.; BÜHRER-SÉKULA, S. Serology: recent developments, strengths, limitations and prospects: a state of the art overview. **Leprosy review**, v. 74, n. 3, p. 196–205, 2003.

PENNA, M. L. F. et al. Comportamento epidemiológico da hanseníase no Brasil. **Secretaria de Vigilância em Saúde, Ministério da Saúde, organizador. Saúde Brasil**, 2009.

QUEIROZ, J. W. et al. Geographic information systems and applied spatial statistics are efficient tools to study Hansen's disease (leprosy) and to determine areas of greater risk of disease. **The American journal of tropical medicine and hygiene**, v. 82, n. 2, p. 306–314, 2010.

RICHARDUS, J. H. Risk factors for the development of clinical leprosy among contacts, and their relevance for targeted interventions. **Lepr rev**, v. 75, p. 310–326, 2004.  
SECRETARIA DE VIGILÂNCIA EM SAÚDE. **Boletim Epidemiológico - Hanseníase**. [s.l: s.n.], v. 49

SHEATHER, S. J.; JONES, M. C. A reliable data-based bandwidth selection method for kernel density estimation. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 53, n. 3, p. 683–690, 1991.

SILVERMAN, B. W. Spline smoothing: the equivalent variable kernel method. **The Annals of Statistics**, p. 898–916, 1984.

SILVERMAN, B. W. Monographs on Statistics and Applied Probability. In: **Density Estimation for Statistics and Data Analysis**. [s.l.] Chapman & Hall, 1986. v. 26.

SNOW, J. **Sobre a maneira de transmissão do cólera**. [s.l.] Programa de Publicações Didáticas, Agência Norte-Americana para o Desenvolvimento Internacional - USAID, 1967.

SPENCER, J. S. et al. Identification of serological biomarkers of infection, disease progression and treatment efficacy for leprosy. **Memórias do Instituto Oswaldo Cruz**, v. 107, p. 79–89, 2012.

TURNBULL, B. W. et al. **Monitoring for clusters of disease: application to leukemia incidence in upstate New York** **Am J Epidemiol**, 1990. Disponível em: [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=2356825](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2356825)

WALKER, S. L.; LOCKWOOD, D. N. J. The clinical and immunological features of

leprosy. **British medical bulletin**, v. 77, n. 1, p. 103–121, 2006.

WALLER, L. A.; GOTWAY, C. A. **Applied spatial statistics for public health data**. [s.l.] John Wiley & Sons, 2004. v. 368

WHO. Global leprosy update, 2017: reducing the disease burden due to leprosy–  
Situation de la lèpre dans le monde, 2017: réduction de la charge de morbidité due à la  
lèpre. **Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire**, v.  
93, n. 35, p. 445–456, 2018.