

**COMPARAÇÃO DAS TÉCNICAS KERNEL E KRIGAGEM
INDICATIVA NA PREDIÇÃO DE VALORES DE VARIÁVEIS
ESPACIALMENTE DISTRIBUÍDAS – ESTUDOS DE CASO**

Juliana Aparecida Gualberto

Dissertação apresentada à Universidade Estadual Paulista “Júlio de Mesquita Filho” para a obtenção do título de Mestre em Biometria.

BOTUCATU
São Paulo - Brasil
Fevereiro – 2020

**COMPARAÇÃO DAS TÉCNICAS KERNEL E KRIGAGEM
INDICATIVA NA PREDIÇÃO DE VALORES DE VARIÁVEIS
ESPACIALMENTE DISTRIBUÍDAS – ESTUDOS DE CASO**

Juliana Aparecida Gualberto

Orientador: Prof. Dr. **José Silvio Govone**

Dissertação apresentada à Universidade Estadual Paulista “Júlio de Mesquita Filho” para a obtenção do título de Mestre em Biometria.

BOTUCATU
São Paulo - Brasil
Fevereiro – 2020

Ficha Catalográfica

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Gualberto, Juliana Aparecida.

Comparação das técnicas Kernel e Krigagem indicativa na
predição de valores de variáveis espacialmente
distribuídas : estudos de caso / Juliana Aparecida
Gualberto. - Botucatu, 2020

Dissertação (mestrado) - Universidade Estadual Paulista
"Júlio de Mesquita Filho", Instituto de Biociências de
Botucatu

Orientador: José Silvio Govone

Capes: 10203001

1. Kernel, Funções de. 2. Krigagem. 3. Doenças
transmissíveis. 4. Dengue. 5. Malária. 6. Análise espacial
(Estatística).

Palavras-chave: Critérios de informação; Dengue; Malária;
Raio de influência; Semivariograma.

ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE MESTRADO DE JULIANA APARECIDA GUALBERTO, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA, DO INSTITUTO DE BIOCIÊNCIAS - CÂMPUS DE BOTUCATU.

Aos 28 dias do mês de fevereiro do ano de 2020, às 09:00 horas, no(a) Sala 3 do Departamento de Bioestatística, reuniu-se a Comissão Examinadora da Defesa Pública, composta pelos seguintes membros: Prof. Dr. JOSÉ SILVIO GOVONE - Orientador(a) do(a) Departamento de Estatística Matemática Aplicada e Computação / Instituto de Geociências e Ciências Exatas - UNESP, Profa. Dra. LICIANA VAZ DE ARRUDA SILVEIRA do(a) Departamento de Bioestatística / Instituto de Biociências de Botucatu - UNESP, Prof. Dr. ENIO JUNIOR SEIDEL do(a) Departamento de Estatística / Universidade Federal de Santa Maria (por meio de videoconferência), sob a presidência do primeiro, a fim de proceder a arguição pública da DISSERTAÇÃO DE MESTRADO de JULIANA APARECIDA GUALBERTO, intitulada **Comparação das Técnicas Kernel e Krigagem Indicativa na Predição de Valores de Variáveis Espacialmente Distribuídas – Estudos de Caso**. Após a exposição, a discente foi arguida oralmente pelos membros da Comissão Examinadora, tendo recebido o conceito final: Aprovada. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelos membros da Comissão Examinadora.

Prof. Dr. JOSÉ SILVIO GOVONE

J. Govone

Profa. Dra. LICIANA VAZ DE ARRUDA SILVEIRA

Liciane Vaz de Arruda Silveira

Prof. Dr. ENIO JUNIOR SEIDEL

Dedicatória

Aos meus pais, Luiz e Maria, que muitas vezes se doaram e renunciaram aos seus sonhos, para que eu pudesse realizar os meus. Quero dizer que essa conquista não é só minha, mas nossa. Tudo que consegui só foi possível graças ao amor, apoio e dedicação que vocês sempre tiveram por mim. Sempre me ensinaram agir com respeito, simplicidade, dignidade, honestidade e amor ao próximo. E graças à nossa união, os obstáculos foram ultrapassados, vitórias foram conquistadas e alegrias divididas.

Agradeço pela paciência e compreensão com minha ausência durante essa longa jornada.

Muitíssimo obrigada.

Agradecimentos

Agradeço à Deus, por ter abençoado todos os dias da minha vida, por iluminar meu caminho e me dar forças para seguir sempre em frente.

Ao professor José Silvio Govone, a oportunidade de tê-lo como orientador de Iniciação Científica e Mestrado. Agradeço pela confiança, pela amizade, conselhos e paciência. O senhor é um exemplo de simplicidade, compreensão e competência.

Aos meus pais Luiz e Maria, pelo apoio, torcida e confiança que sempre depositam em mim; pelos momentos que não estivemos juntos e souberam entender.

Aos meus amigos: Bethina, Eduardo, Elizabete, Gabriela, Gustavo e Marta, pelo privilégio de ter essas amizades, pelo apoio, incentivo, pelos bons momentos, por terem tornado o dia a dia na pós-graduação tão prazeroso! Foi extremamente enriquecedor conhecer e conviver com cada um de vocês.

Aos professores, funcionários e colegas do Programa de Pós Graduação em Biometria – Universidade Estadual Paulista “Julio de Mesquita Filho” – Botucatu, pela convivência gratificante durante todo esse percurso.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES pelo apoio financeiro.

Às demais pessoas que contribuíram direta ou indiretamente na elaboração deste trabalho ou participaram da minha vida, e que, porventura, eu tenha me esquecido de agradecer.

Sumário

	Página
LISTA DE FIGURAS	ix
LISTA DE TABELAS	xiii
RESUMO	xiv
SUMMARY	xvi
1 INTRODUÇÃO	1
1.1 Objetivos	2
1.1.1 Geral:	2
1.1.2 Particulares:	2
1.2 Dengue	3
1.3 Malária	4
2 REVISÃO DE LITERATURA	6
2.1 Análise espacial de dados	6
2.2 Geoestatística	9
2.3 Krigagem	9
2.4 O variograma	10
2.4.1 Propriedades do variograma	13
2.4.2 Modelos Variográficos	15
2.4.2.1 Modelos com Patamar	16
2.4.2.2 Modelos sem Patamar	18

2.5	Tipos de Krigagem	19
2.6	Krigagem Ordinária	19
2.7	Krigagem Indicativa	22
2.8	Análise da dependência espacial	23
2.9	Processos Pontuais	25
2.10	Estimador Kernel	26
2.11	Método do Vizinho Mais Próximo	30
3	MATERIAL E MÉTODOS	35
3.1	Material	35
3.1.1	Banco - Dengue	35
3.1.2	Banco - Malária	36
3.2	Métodos	37
3.2.1	Krigagem Indicativa para os dados de dengue	38
3.2.2	Kernel para os dados de dengue	39
3.2.3	Krigagem Indicativa para os dados de malária	40
3.2.4	Kernel para os dados de malária	41
4	RESULTADOS	42
4.1	Análise exploratória dos dados	42
4.2	Análise exploratória espacial dos dados	43
4.2.1	Análise exploratória espacial dos dados de dengue	43
4.2.2	Análise exploratória espacial dos dados de malária	46
4.3	Semivariograma dados de dengue	48
4.4	Krigagem Indicativa para os dados de dengue	52
4.5	Ajuste da função K de Ripley para os dados de dengue	54
4.6	Kernel para os dados de dengue	56
4.7	Comparação das técnicas Kernel e Krigagem Indicativa para os dados de dengue	58
4.8	Semivariograma dados de malária	59

	viii
4.9 Krigagem Indicativa para os dados de malária	64
4.10 Ajuste da função K de Ripley para os dados de malária	66
4.11 Kernel para os dados de malária	67
4.12 Comparação das técnicas Kernel e Krigagem Indicativa para os dados de malária	69
5 CONSIDERAÇÕES FINAIS	71
REFERÊNCIAS BIBLIOGRÁFICAS	73

Lista de Figuras

	Página
1 Mapa da população indígena total em 2010 no Brasil (Instituto Brasileiro de Geografia e Estatística–IBGE, 2019).	7
2 Mapa batimétrico de águas rasas no Reino do Bahrein no Golfo Árabe (Bannari & Kadhem, 2017).	8
3 Mapa da distribuição mundial e países com as maiores taxas de incidência de Sarampo (02/2018 – 01/2019) (Ministério da Saúde, 2019b).	8
4 Comparativo da correlação espacial entre amostras em diferentes regiões de um variograma (Pereira, 2017).	14
5 Modelo teórico do ajuste do variograma (Wanderley et al., 2012).	14
6 Representação gráfica de modelos com patamar (Monteiro et al., 2004a).	16
7 Representação gráfica de modelos sem patamar (Monteiro et al., 2004a).	18
8 Ocorrência espacialmente aleatória, aglomerada e regular (Andrade et al., 2007).	25
9 Estimador de intensidade de distribuição de pontos (Monteiro et al., 2004b).	27
10 Passos para o cálculo de desenvolvimento de densidade de pontos segundo a técnica de Kernel (Andrade et al., 2007).	28
11 Gráfico de $\hat{G}(w)$ em relação as distâncias (Andrade et al., 2007).	31
12 O gráfico à esquerda representa tendências para agrupamento e o gráfico à direita representa padrões de regularidade. Os traçados no gráfico estarão acima ou abaixo da linha de 45° , sendo a linha verde o envelope inferior, a linha azul o estimado e a linha vermelha o envelope superior (Vasconcelos, 2016).	32

13	Gráfico e interpretação da plotagem de $\hat{L}(h)$ (Andrade et al., 2007).	33
14	Análise do gráfico $\hat{L}(h)$ com os envelopes (Andrade et al., 2007).	34
15	Mapa da área urbana da cidade de Rio Claro-SP com os 197 setores censitários. Fonte: Aplicativo Google Earth Pro.	36
16	Mapa da área urbana da cidade de Porto Velho-RO com os 90 setores censitários. Fonte: Aplicativo Google Earth Pro.	37
17	Histograma de frequência: à direita, da contagem de casos de dengue confirmados no ano de 2011 em Rio Claro-SP e à esquerda, da contagem de casos de malária confirmados no ano de 1999 em Porto Velho-RO. . .	43
18	Distribuição espacial dos setores censitários da cidade de Rio Claro-SP. .	44
19	Análise de tendência casos de dengue confirmados no ano de 2011 em Rio Claro-SP. O primeiro gráfico de quantis está identificado de acordo com o quantil de cada setor censitário. O segundo e terceiro são gráficos de dispersão da contagem de casos e o quarto é o histograma referente à densidade da contagem de casos.	45
20	Gráfico boxplot da contagem de casos de dengue confirmados no ano de 2011 em Rio Claro-SP.	45
21	Distribuição espacial dos setores censitários da cidade de Porto Velho-RO.	46
22	Análise de tendência casos de malária confirmados no ano de 1999 em Porto Velho-RO. O primeiro gráfico de quantis está identificado de acordo com o quantil de cada setor censitário. O segundo e terceiro são gráficos de dispersão da contagem de casos e o quarto é o histograma referente a densidade da contagem de casos.	47
23	Gráfico boxplot da contagem de casos de malária confirmados no ano de 1999 em Porto Velho-RO.	47
24	Semivariogramas dos dados binários de dengue, com as distâncias máximas: (a) distância máxima, (b) metade da distância máxima, (c) 1/3 da distância máxima e (d) 1/4 da distância máxima, respectivamente.	48
25	Gráfico da anisotropia dos dados binários dos casos de dengue.	49

26	Gráficos dos ajustes dos semivariogramas: à esquerda com Mínimos Quadrados Ordinários (OLS) e à direita com Mínimos Quadrados Ponderados (WLS), para os casos de dengue confirmados no ano de 2011 em Rio Claro-SP.	50
27	Gráficos dos ajustes dos semivariogramas: à esquerda com a Máxima Verossimilhança (ML) e à direita com a Máxima Verossimilhança Restrita (REML), para os casos de dengue confirmados no ano de 2011 em Rio Claro-SP.	51
28	Mapa de isolinhas dos casos de dengue confirmados no ano de 2011 em Rio Claro-SP.	53
29	Mapa de Krigagem Indicativa referente aos casos de dengue confirmados no ano de 2011 em Rio Claro-SP.	53
30	Variância de Krigagem referente aos casos de dengue confirmados no ano de 2011 em Rio Claro-SP.	54
31	Gráfico da função K (função de Ripley) dos casos de dengue confirmados no ano de 2011 em Rio Claro-SP.	55
32	Gráfico da função F (função de espaço vazio ou taxa de risco) dos casos de dengue confirmados no ano de 2011 em Rio Claro-SP.	55
33	Gráfico do raio de influência, com $\sigma = 241,0809$, para os casos de dengue confirmados no ano de 2011 em Rio Claro-SP.	56
34	Mapa de estimação da densidade Kernel referente aos casos de dengue confirmados no ano de 2011 em Rio Claro-SP.	57
35	Mapa de incerteza da estimação da densidade Kernel referente aos casos de dengue confirmados no ano de 2011 em Rio Claro-SP.	57
36	Comparação dos mapas para os dados de dengue, à direita o mapa Kernel e à esquerda o mapa da Krigagem Indicativa.	59
37	Semivariogramas dos dados binários de malária, com as: (a) distância máxima, (b) metade da distância máxima, (c) 1/3 da distância máxima e (d) 1/4 da distância máxima, respectivamente.	60

38	Gráfico da anisotropia dos dados binários dos casos de malária.	61
39	Gráficos dos ajustes dos semivariogramas: à esquerda com Mínimos Quadrados Ordinários (OLS) e à direita com Mínimos Quadrados Ponderados (WLS), para os casos de malária confirmados no ano de 1999 em Porto Velho-RO.	62
40	Gráficos dos ajustes dos semivariogramas: à esquerda com a Máxima Verossimilhança (ML) e à direita com a Máxima Verossimilhança Restrita (REML), para os casos de malária confirmados no ano de 1999 em Porto Velho-RO.	62
41	Mapa de isolinhas dos casos de malária confirmados no ano de 1999 em Porto Velho-RO.	64
42	Mapa de Krigagem Indicativa referente aos casos de malária confirmados no ano de 1999 em Porto Velho-RO.	65
43	Variância de Krigagem referente aos casos de malária confirmados no ano de 1999 em Porto Velho-RO.	65
44	Gráfico da função K (função de Ripley) dos casos de malária confirmados no ano de 1999 em Porto Velho-RO.	66
45	Gráfico da função F (função de espaço vazio ou taxa de risco) dos casos de malária confirmados no ano de 1999 em Porto Velho-RO.	67
46	Gráfico do raio de influência, com $\sigma = 208,2787$, para os casos de malária confirmados no ano de 1999 em Porto Velho-RO.	68
47	Mapa de estimação da densidade Kernel referente aos casos de malária confirmados no ano de 1999 em Porto Velho-RO.	68
48	Mapa de incerteza da estimação da densidade Kernel referente aos casos de malária confirmados no ano de 1999 em Porto Velho-RO.	69
49	Comparação dos mapas para os dados de malária, á direita o mapa Kernel e á esquerda o mapa da Krigagem Indicativa.	70

Lista de Tabelas

Página

1	Resultados das análises exploratórias, por setores, das contagens de casos de dengue confirmados no ano de 2011 em Rio Claro-SP e das contagens de casos de malária no ano de 1999 em Porto Velho-RO.	42
2	Autovalidação dos ajustes Mínimos Quadrados Ordinários (OLS) e Mínimos Quadrados Ponderados (WLS) para os dados de dengue.	51
3	Resultados referentes a modelagem do semivariograma para à Krigagem Indicativa dos casos de dengue.	52
4	Variância de Krigagem Indicativa dos casos de dengue.	54
5	Critérios de informação Akaike (AIC) e Bayesiano (BIC) da estimação de densidade Kernel dos casos de dengue.	58
6	Critérios de informação Akaike (AIC) e Bayesiano (BIC) da estimação de densidade Kernel e da Krigagem Indicativa dos casos de dengue.	59
7	Autovalidação dos ajustes Mínimos Quadrados Ordinários (OLS) e Mínimos Quadrados Ponderados (WLS) para os dados de malária.	63
8	Resultados referentes a modelagem do semivariograma para a Krigagem Indicativa dos casos de malária.	63
9	Variância de Krigagem Indicativa dos casos de malária.	66
10	Critérios de informação Akaike (AIC) e Bayesiano (BIC) da estimação de densidade Kernel dos casos de malária.	69
11	Critérios de informação Akaike (AIC) e Bayesiano (BIC) da estimação de densidade Kernel e da Krigagem Indicativa dos casos de malária.	70

COMPARAÇÃO DAS TÉCNICAS KERNEL E KRIGAGEM INDICATIVA NA PREDIÇÃO DE VALORES DE VARIÁVEIS ESPACIALMENTE DISTRIBUÍDAS – ESTUDOS DE CASO

Autora: JULIANA APARECIDA GUALBERTO

Orientador: Prof. Dr. JOSÉ SILVIO GOVONE

RESUMO

O Kernel e a Krigagem Indicativa são duas técnicas utilizadas na análise espacial para estimar a ocorrência de determinado fenômeno em uma área de estudo. Com o conhecimento da distribuição espacial para casos de doenças epidêmicas é possível criar estratégias de controle de transmissão. O objetivo deste trabalho foi estudar ambas técnicas de estatística espacial usando conjuntos de dados epidemiológicos, dengue e malária, de duas cidades, Rio Claro-SP e Porto Velho-RO, respectivamente, a fim de observar os locais com maiores ocorrências e exemplificar para comparação das técnicas para verificar qual teve melhor precisão nas previsões. A partir desses resultados é possível verificar as áreas em que ocorreram os casos e tomar providências para diminuição ou erradicação das mesmas. A técnica Krigagem Indicativa gerou mapas que representam a probabilidade do valor de corte

a ser superado e a técnica kernel gerou mapas que representam a intensidade de ocorrência em cada ponto. Ambas foram eficientes ao mostrar as áreas de concentração das endemias, mas a falta de variáveis adicionais nos bancos de dados, nos impôs algumas limitações ao fazer as comparações das técnicas; com isso, o melhor meio de compará-las foi através dos critérios de informação Akaike (AIC) e Bayesiano (BIC). Os valores encontrados nos critérios de informação para ambas as técnicas foram condizentes com os encontrados na literatura, mas para uma comparação não foi tão eficaz, pois apresentou valores muito distantes, porém com um critério visual dos mapas, umas das técnicas se destacou para os dois conjuntos de dados. Ao final concluímos que essas técnicas não são comparáveis, mas sim complementares, pois a técnica Kernel mostrou a intensidade das endemias e a técnica Krigagem Indicativa mostrou a probabilidade desses valores serem maiores que um ponto de corte determinado.

Palavras-chave: Critérios de Informação. Semivariograma. Raio de Influência. Dengue. Malária. Rio Claro. Porto Velho.

COMPARISON OF KERNEL AND INDICATIVE KRIGING TECHNIQUES IN PREDICTING VALUES OF SPATIALLY DISTRIBUTED VARIABLES – CASE STUDIES

Author: JULIANA APARECIDA GUALBERTO

Adviser: Prof. Dr. JOSÉ SILVIO GOVONE

SUMMARY

The Kernel and the Indicative Kriging are two techniques used in spatial analysis to estimate the occurrence of a certain phenomenon on a field of study. With knowledge on spatial distribution for cases of epidemic diseases, it is possible to create transmission control strategies. The objective of this paper was to study both spatial statistics techniques using epidemiological data sets, dengue and malaria, from two cities, Rio Claro-SP and Porto Velho-RO, respectively, in order to observe the places which present the highest occurrence rates and to exemplify comparing the techniques and to observe which ones presented a better accuracy on their predictions. From these results it is possible to check the areas where the cases occur and take steps to reduce or eradicate them. The Indicative Kriging technique generated maps that represent a probability of the cut value being surpassed and the

Kernel technique generated maps that represent the intensity of occurrence at each point. Both of them were efficient in showing areas of concentration of endemics, but the lack of additional variables in the databases ended up causing some restrictions when making comparing the techniques; therefore, the best means of comparison was through the use of Akaike (AIC) and Bayesian (BIC) information. The values found in the information criteria for both techniques were similar to those found in literature, but as a means of comparison it was not as effective, as it shows very distant values, but with a visual criteria of the maps, one of the techniques was highlighted for the two data sets. In the end, we conclude that these techniques are not comparable, but rather complementary, since a kernel technique showed intensity of endemic diseases and an Indicative Kriging technique showed the probability of these values being greater than a determined cutoff point.

Keywords: Information Criteria. Semivariogram. Bandwidth. Dengue. Malaria. Rio Claro. Porto Velho.

1 INTRODUÇÃO

O estudo da distribuição espacial de fenômenos tem sido cada vez mais comum em diversas áreas do conhecimento (como por exemplo, na saúde, agronomia, geologia). Isso se deve a capacidade da análise espacial de modelar dados de diferentes formatos e obter informações georeferenciadas distribuídas numa área, além de gerar modelos de explicação, a fim de prever o comportamento da variável e ressaltar suas características diretamente ligadas a localização espacial, proporcionando resultados mais significativos (Barbosa, 2014).

Dentre as abordagens utilizadas na análise espacial estão a técnica Kernel, que consiste em um método não paramétrico que estima a função de densidade de probabilidade de uma variável aleatória, cujo objetivo é suavizar a superfície, calculando a densidade em qualquer ponto da região, com base nos valores da variável na vizinhança dos pontos através da interpolação, preservando as características e variabilidade da variável (Diggle & Marron, 1988); e a técnica Krigagem Indicativa, que consiste em um método não linear que modela a incerteza sobre o valor desconhecido, a partir da função de distribuição acumulada condicional, daquela variável (Journel, 1983).

Ambas técnicas são aplicadas para estimar as chances de determinado fenômeno estar ocorrendo em pontos de uma área de estudo e, um comparativo delas, em alguns conjuntos de dados, permite verificar qual tem a melhor precisão nas predições.

Um dos conjuntos de dados no qual é possível fazer essa comparação das técnicas são dados de doenças endêmicas.

As doenças endêmicas podem ser entendidas como a ocorrência de um

agravo dentro de um número esperado de casos para certa região, em determinado período de tempo, baseado na sua ocorrência em anos anteriores não epidêmicos. Desta forma, a incidência de uma doença endêmica é relativamente constante, podendo ocorrer variações sazonais no comportamento esperado para o agravo em questão (Moura & Rocha, 2012).

No Brasil, pouco se sabe sobre a distribuição espacial das doenças endêmicas em áreas urbanas. Com os deslocamentos populacionais das áreas rurais para áreas urbanas, houve o fenômeno de expansão das metrópoles, tornando mais complexo o controle da transmissão de algumas endemias, exigindo novas estratégias de controle (Hino et al., 2006).

Existem várias endemias que desafiam a saúde pública brasileira e, dentre as principais estão a dengue e a malária (Dias, 1998). Entretanto, com o conhecimento da distribuição espacial das ocorrências é possível localizar as áreas com maior concentração e criar planejamentos de políticas públicas para o controle, diminuição ou erradicação das endemias.

1.1 Objetivos

1.1.1 Geral:

Dispondo dessas informações, o objetivo geral deste trabalho foi estudar e comparar as técnicas de estatística espacial, Kernel e Krigagem Indicativa, para os conjuntos de dados epidemiológicos de dengue e malária.

1.1.2 Particulares:

- Aplicar as técnicas Kernel e Krigagem Indicativa para um mesmo conjunto de dados;
- Calcular os critérios de informação Akaike (AIC) e Bayesiano (BIC);
- Comparar as técnicas para verificar qual apresenta melhor precisão nas predições.

Os tópicos 1.2 e 1.3 abordam sobre as doenças, dengue e malária, para uma melhor compreensão da aplicação com dados epidemiológicos.

1.2 Dengue

Os vírus da dengue (DEN) são arbovírus, encontrados em áreas tropicais e subtropicais, deixando cerca de 3 bilhões de pessoas em regiões sob o risco de infecção (Souza, 2008).

No Brasil, a região Sudeste é a que registra o maior número de casos de dengue por ano, devido suas condições climáticas e o ambiente sinantrópico. Em seguida, por ordem de incidência são Nordeste, Centro-Oeste, Sul e Norte (Costa et al., 2013).

A dengue é uma doença viral de transmissão vetorial, causada por um dos quatro sorotipos do vírus dengue (DENV-1 a DENV-4), pertencentes ao gênero *Flavivírus*, da família *Flaviviridae* (Costa et al., 2013).

Após a infecção do vírus através do seu vetor, o mosquito *Aedes aegypti*, o período de incubação será de 3 a 15 dias, em seguida, a doença poderá evoluir para outras formas clínicas de acordo com a Organização Mundial de Saúde (OMS): assintomática, indiferenciada, dengue clássica e febre hemorrágica por dengue (Souza, 2008).

Sem tratamento específico e sem uma vacina eficaz disponível contra a dengue, uma das formas de prevenção existente consiste no controle do mosquito transmissor (Costa et al., 2013). Campanhas informativas, que utilizam redes de televisão, rádios, jornais, folhetos, entre outros, buscam a colaboração da população para a eliminação dos focos de mosquitos, porém tem eficiência limitada, por isso são necessárias ações ambientais e da vigilância epidemiológica, entomológica e viral (Claro et al., 2004).

No estado de São Paulo, a infestação pelo vetor ocorreu em 1985, devido as condições ideais para a proliferação do *Aedes aegypti* e posteriormente, em 1990, ocorreu à circulação do vírus da dengue (Piovezan, 2009).

1.3 Malária

A malária é uma doença infecciosa febril aguda e, segundo a Organização Mundial da Saúde (OMS), é uma doença essencialmente tropical e subtropical, com prevalência nas regiões e países inseridos dentro dos trópicos úmidos (Ferreira, 2015).

Em 2015, cerca de 3,2 bilhões de pessoas estavam em regiões com risco de contrair malária. A maioria dos casos e mortes relacionadas à doença ocorreram na África Subsaariana, mas há casos registrados na Ásia, América Latina e em menor proporção, no Oriente Médio (OPAS, 2016).

O vírus é causado por um dos quatro protozoários do gênero *Plasmodium*: *Plasmodium vivax*, *Plasmodium falciparum*, *Plasmodium malariae* e *Plasmodium ovale*, transmitidos pela fêmea infectada do mosquito *Anopheles*. No Brasil, os três primeiros protozoários são os mais presentes (Governo do Estado de São Paulo - Secretaria de Estado da Saúde, 2019).

A maioria dos casos de malária no Brasil, se concentra na região Amazônica, nos estados do Acre, Amapá, Amazonas, Maranhão, Mato Grosso, Pará, Rondônia, Roraima e Tocantins, devido ao fácil contato homem-mosquito. Nas demais regiões, apesar das poucas notificações, a doença não pode ser negligenciada, pois se observa uma letalidade mais elevada do que na região Amazônica (Ministério da Saúde, 2019a).

Após a realização da campanha de erradicação, durante a década de 60, o número de casos de malária atingiu o seu valor mais baixo, 52.469 casos, confinando-se a transmissão, praticamente, à região amazônica (Barata, 1995).

A malária não é uma doença contagiosa, mas com a movimentação dos indivíduos por todo o território, pode reintroduzir a transmissão através do seu vetor, em regiões onde ela já havia sido controlada (Barata, 1995).

Nos anos 80 e 90 com a urbanização da região amazônica, devido a construção de estradas, usinas hidroelétricas, o desenvolvimento agropecuário e as atividades de mineração, ocorreu um aumento considerável na transmissão (Barata,

1995).

O diagnóstico correto da infecção malárica só é possível pela demonstração do parasito, ou de antígenos relacionados, no sangue periférico do paciente. Com a confirmação da malária, o paciente recebe o tratamento em regime ambulatorial, com comprimidos que são fornecidos gratuitamente em unidades do Sistema Único de Saúde (SUS). Somente os casos graves são hospitalizados de imediato (Ministério da Saúde, 2019a).

Existem algumas medidas de prevenção, como o uso de mosquiteiros, pequenas obras de saneamento para eliminação de criadouros do vetor, limpeza das margens dos criadouros, entre outros (Ministério da Saúde, 2019a).

2 REVISÃO DE LITERATURA

2.1 Análise espacial de dados

Segundo Monteiro et al. (2004c), a ênfase da análise espacial é mensurar propriedades e relacionamentos, levando em conta a localização espacial do fenômeno em estudo de forma explícita. Ou seja, a ideia central é incorporar o espaço à análise que se deseja fazer.

De acordo com Bailey & Gatrell (1995), a taxonomia mais utilizada para caracterizar os problemas em análise espacial considera três tipos de dados, os Eventos ou Padrões Pontuais, as Superfícies Contínuas e as Áreas com Contagens e Taxas Agregadas, que são definidos da seguinte forma:

- *Eventos ou Padrões Pontuais*: Consistem em uma série de pontos em que seus fenômenos são expressos através de ocorrências localizadas em uma área de interesse, exemplo Figura 1.

- *Superfícies Contínuas:* São estimadas por conjunto de amostras de campo resultantes de levantamento de recursos naturais, e podem estar regularmente ou irregularmente distribuídos, exemplo Figura 2.

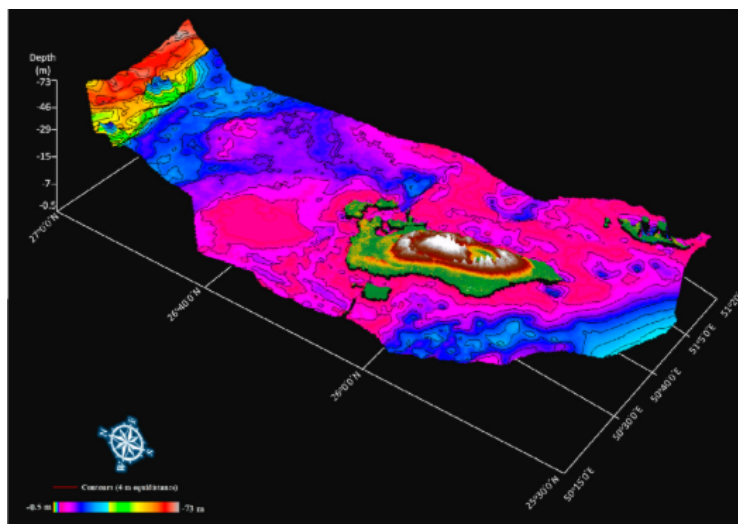


Figura 2: Mapa batimétrico de águas rasas no Reino do Bahrein no Golfo Árabe (Bannari & Kadhem, 2017).

- *Áreas com Contagens e Taxas Agregadas:* Trata-se de dados associados a levantamentos populacionais e estão disponíveis em áreas delimitadas por polígonos, exemplo Figura 3.

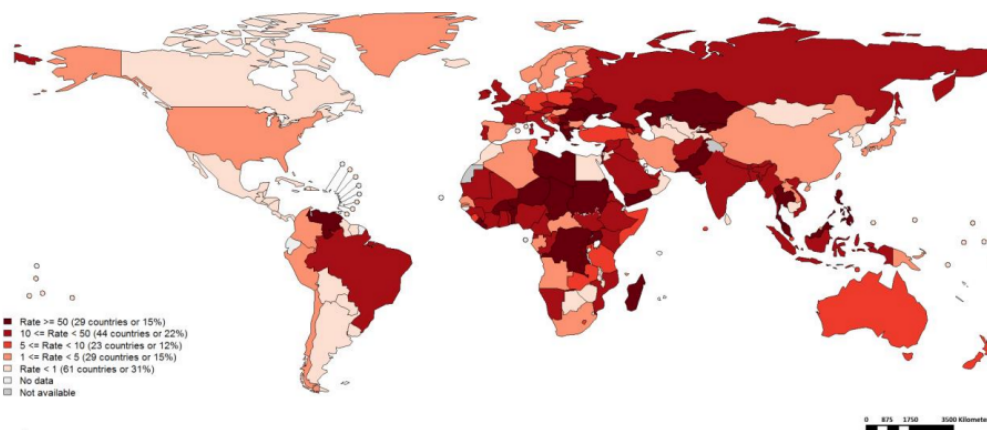


Figura 3: Mapa da distribuição mundial e países com as maiores taxas de incidência de Sarampo (02/2018 – 01/2019) (Ministério da Saúde, 2019b).

2.2 Geoestatística

A geoestatística surgiu no início dos anos sessenta quando o engenheiro de mineração sul-africano Daniel Krige, que trabalhava com dados de concentração de ouro, não conseguiu encontrar sentido nas variâncias se não considera-se a distância entre as amostras, então o professor Georges Matheron, inspirado nessas observações e nos trabalhos de H. J. de Wijs, professor da Universidade Técnica de Delf, na Holanda, desenvolveu uma teoria chamada Teoria das Variáveis Regionalizadas, que contém os fundamentos da geoestatística (Santos, 2015).

Segundo Matheron (1963), a variável regionalizada é definida como uma função espacial numérica, que varia de um local para outro, com uma continuidade aparente, ou seja, as variáveis dependem de um valor e da sua posição espacial. Essa variação não pode ser representada por uma função matemática simples, então a continuidade ou dependência espacial pode ser estimada pela função semivariograma.

Na geoestatística existe um método de interpolação chamado krigagem, no qual, o nome dado por Matheron (1963) é em homenagem ao engenheiro Daniel Krige.

2.3 Krigagem

A palavra Krigagem, tradução do francês *Krigeage* e do inglês *Kriging* é um método de interpolação, que usa a dependência espacial entre amostras vizinhas, para estimar valores em qualquer posição em uma área de interesse, sem tendência e com variância mínima.

Atualmente, com a necessidade de obter melhores previsões a partir de dados observados é aplicada em diferentes campos da ciência, como por exemplo a física (Avaliação de Incerteza para Propagação de Som ao Ar Livre)(Leroy et al., 2010), a biologia (Distribuição Espacial de Espécies Arbóreas em uma Floresta Estacional em Santa Maria, Sul do Brasil)(Giehl et al., 2007) e a medicina (Distribuição

Espacial do Índice de Desenvolvimento Humano, da Infecção pelo HIV e da Comorbidade AIDS-tuberculose: Brasil, 1982 – 2007)(Rodrigues Jr et al., 2014).

Símile à vários métodos que usam um conjunto limitado de pontos distribuídos no espaço para estimar uma variável de interesse, o que difere a krigagem é a maneira que o método atribui pesos às diferentes amostras, por exemplo, no caso de interpolação linear simples os pesos são todos iguais a $1/N$ (N = número de amostras) (Camargo, 1997); na interpolação baseada no inverso do quadrado das distâncias, os pesos são definidos como o inverso do quadrado da distância que separa o valor interpolado dos valores observado, enquanto na Krigagem, o método é semelhante ao de interpolação por média móvel ponderada, porém neste caso os pesos são determinados a partir de uma análise espacial, baseada no variograma experimental.

A Krigagem tem a vantagem de prever estimativas da incerteza em torno de cada valor interpolado, também permite usar dados coletados usando diferentes métodos de amostragem e, além disso, é um método simples, pois é baseado apenas em dados empíricos.

Se houver pouca autocorrelação espacial, a Krigagem não se destaca em relação a outros métodos de interpolação, mas ela pode ser mais útil se a autocorrelação for moderada, pois irá preservar a variação espacial, o que pode não ocorrer com outros métodos.

A Krigagem pode ser entendida como um processo de duas etapas: primeiro, a estrutura de covariância espacial dos pontos amostrados é determinada pelo ajuste de um variograma; segundo, os pesos derivados dessa estrutura de covariância são usados para interpolar valores para pontos ou blocos não amostrados no campo espacial (University Columbia, 2019).

2.4 O variograma

O variograma é uma ferramenta básica de suporte às técnicas de Krigagem, que permite representar quantitativamente a variação de um fenômeno regiona-

lizado no espaço (Journel & Huijbregts, 1978), o mesmo possui duas características, a “teórica” ou “modelo”, que é a distribuição que melhor ajusta aos dados; e a “experimental”, sendo o gráfico dos valores observados.

A função variograma pode ser definida como sendo “a esperança matemática do quadrado da diferença entre os valores de pontos no espaço, separados por uma distância h ” (Yamamoto, 1991). Matematicamente, seja um conjunto de variáveis aleatórias $Z(x)$, onde $Z(x) = Z(x_i)$, $i = 1, \dots, n$, correlacionadas entre si. A média é dada por:

$$m = E[Z(x)] \quad (1)$$

A variância associada à média é calculada como:

$$Var[Z(x)] = E \{ [Z(x) - m]^2 \} \quad (2)$$

A covariância mede o nível de correspondência entre valores da mesma variável obtidos em pontos separados por uma distância h , conforme uma determinada distância, ela pode ser calculada da seguinte maneira:

$$C(h) = E \{ [Z(x+h) - m][Z(x) - m] \} \quad (3)$$

Para distância nula ($h = 0$), temos que a covariância é igual a variância da variável regionalizada $Z(x)$.

Segundo Yamamoto & Landim (2015), a função variograma é definida como a variância do incremento $[Z(x+h) - Z(x)]$:

$$\gamma(h) = \frac{1}{2} E[Z(x) - Z(x+h)]^2 \quad (4)$$

então,

$$2\gamma(h) = E[Z(x) - Z(x+h)]^2 = Var[Z(x) - Z(x+h)]. \quad (5)$$

O prefixo “semi” é devido a Equação (5) do variograma que pode ser escrita como a média das diferenças ao quadrado dividido por dois, portanto, $2\gamma(h)$ é chamando de variograma e $\gamma(h)$ de semivariograma (Yamamoto & Landim, 2015).

O semivariograma é utilizado unicamente em condições estacionárias ou de quase-estacionaridade, pois se o valor médio esperado sofrer mudanças graduais no espaço caracteriza uma tendência (*drift*), logo outros recursos de avaliação estrutural deverão ser aplicados (Sturaro, 2015).

Quando o vetor distância h for infinitamente pequeno, o variograma será mínimo e a covariância máxima, e a medida em que h aumenta, o variograma eleva e a covariância diminui, ocorrendo uma independência maior entre os pontos. Porém existe uma medida de h em que a variograma e a covariância apresentam os mesmos valores.

Pode-se afirmar que uma variável regionalizada não estacionária possui duas componentes significativas:

- A tendência (*drift*), é o valor esperado da variável regionalizada numa vizinhança qualquer: $m = E[Z(x)]$. Este conceito é semelhante ao de valor ajustado em superfícies de tendência.
- O resíduo, constitui a diferença entre o valor real obtido e o valor esperado: $Y(x) = Z(x) - m$.

A importância deste desdobramento, é que os resíduos possuem média zero e são estacionários, o que viabiliza a elaboração dos semivariogramas e perfaz a análise de variabilidade (Sturaro, 2015).

Um conceito básico na teoria das regionalizadas é a hipótese intrínseca, a qual implica que uma função (a função intrínseca) descreve o comportamento espacial da variável regionalizada dentro do espaço e que essa função é uma característica intrínseca da regionalização (Yamamoto, 2001). Esta função intrínseca é conhecida como semivariograma. A hipótese intrínseca, algumas vezes é referenciada como hipótese de quase-estacionaridade, pois a variação espacial é estacionária se o variograma for o mesmo onde quer que se amostra.

2.4.1 Propriedades do variograma

A identificação e a análise do variograma são feitas através de propriedades gráficas da função, as quais são:

Alcance ou Amplitude (a): é a distância dentro da qual os dados são correlacionados. Tal valor representa a quantificação do conceito de zona de influência de uma amostra, de tal forma que teores separados por distâncias maiores que o alcance da zona de influência não estão correlacionados (Matheron, 1963). Segundo Yamamoto (1991), “é a distância a partir da qual as amostras passam ser independentes”. Para Matheron (1963), o parâmetro é a distância a partir da qual o variograma atinge o seu patamar. Deste ponto em diante, considera-se que não existe mais dependência espacial entre as amostras, porque a variância da diferença entre pares de amostras $Z(x)$ e $Z(x + h)$ torna-se invariante com a distância (Monteiro et al., 2004a);

Patamar (C): é o valor no qual o variograma se estabiliza no campo aleatório (para casos estacionários) e é numericamente igual à variância da variável regionalizada (Yamamoto, 1991), conforme o exemplo da Figura 4.

O Efeito Pepita (C_0): é o valor do variograma na origem ($h = 0$). Entretanto, o valor na origem deveria ser nulo, pois duas amostras tomadas no mesmo ponto ($h = 0$) deveriam ter os mesmos valores (Yamamoto, 1991), mas à medida que h tende para 0 (*zero*), $\gamma(h)$ pode se aproximar de um valor positivo chamado Efeito Pepita, o que caracteriza a variabilidade de escala muito curta. É comum também usar o termo variabilidade de escala curta quando se refere a distâncias de variograma menores que o menor espaçamento entre os pontos de amostragem (Rossi & Deutsch, 2013).

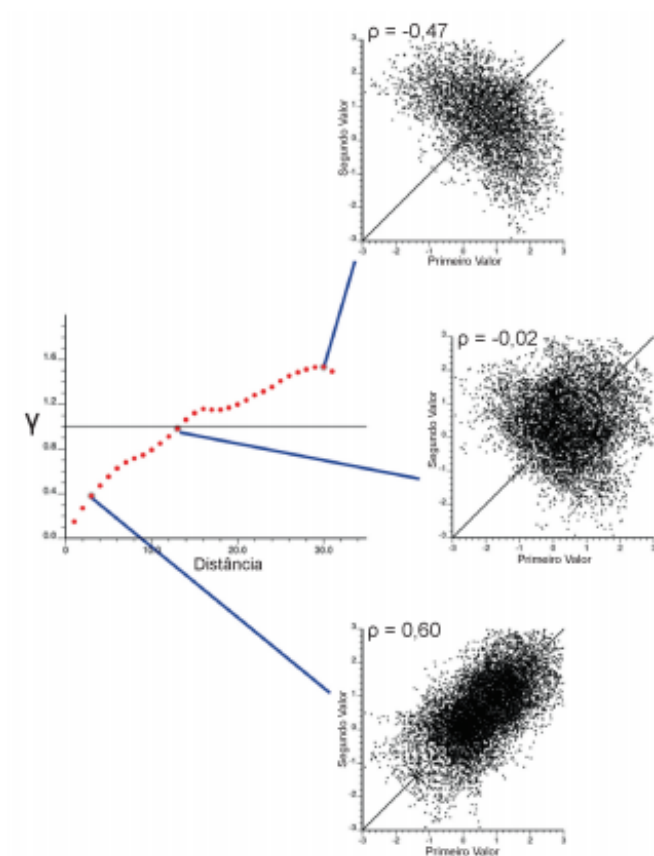


Figura 4: Comparativo da correlação espacial entre amostras em diferentes regiões de um variograma (Pereira, 2017).

A Figura 5 apresenta o modelo teórico de um variograma com patamar:

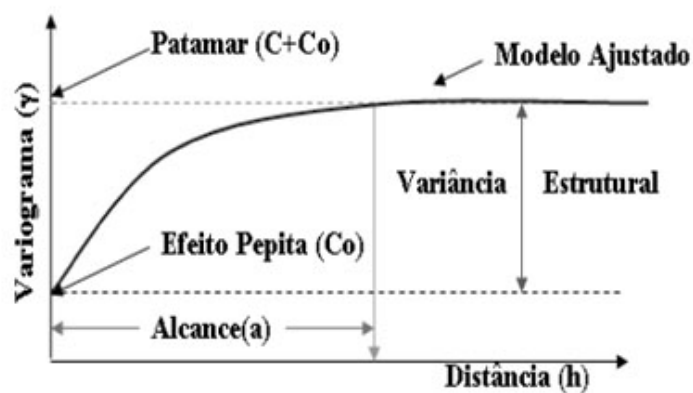


Figura 5: Modelo teórico do ajuste do variograma (Wanderley et al., 2012).

Uma variável regionalizada é estacionária se as estatísticas de primeira, segunda, até graus de maior ordem da variável aleatória forem as mesmas para qualquer vetor de distância h , ou seja, tem a propriedade de que a média, variância e estrutura de autocorrelação não mudam no decorrer do tempo.

Sejam $Z(x)$ e $Z(x + h)$ dois valores de uma variável regionalizada obtidos nos pontos x e $x + h$, separados por uma distância h , vetor com direção e orientação específica no espaço; a diferença entre esses dois valores ($Z(x) - Z(x + h)$) é o valor de outra variável aleatória (Landim, 2003).

Para uma variável estacionária de primeira ordem, é admitido que sua esperança seja um valor m constante em toda área, este parâmetro passa a ser independente da localização de x_i e pode ser estimado pela média aritmética dos valores das realizações das variáveis aleatórias:

$$E[Z(x) - Z(x + h)] = m = 0 \quad (6)$$

em que m representa a tendência ou deriva (*drift*).

Segundo Journel & Huijbregts (1978), a estacionariedade de segunda ordem assume a existência da variância e, portanto, de uma variância a priori finita. Existem porém, fenômenos físicos e, conseqüentemente, variáveis regionalizadas com uma capacidade infinita de dispersão, nos quais não se pode definir, a priori, nem a covariância, nem a variância, mas se pode determinar um variograma.

2.4.2 Modelos Variográficos

Para empregar um modelo variográfico $\gamma(h)$ que melhor se ajusta ao comportamento espacial da variável estudada, são construídas malhas com disposição regular ou irregular, que seguidamente são regularizadas, no qual os cálculos são feitos em direções previamente estabelecidas, objetivando a variabilidade espacial do fenômeno em estudo. Feito os semivariogramas dos valores experimentais, procura-se ajustar um modelo matemático que mais se aproxima da configuração dos mesmos, porém podem existir uma infinidade de funções que ajustem aos semivariogramas

experimentais, mas como a prática tem mostrado, alguns modelos apoiados nas suposições teóricas das variáveis regionalizadas, têm satisfeito a maioria das suas aplicações.

2.4.2.1 Modelos com Patamar

Nestes tipos de modelos normalmente os ajustes representam a estacionaridade de segunda ordem. A função semivariograma cresce na medida que aumenta a distância h entre as amostras, até atingir um patamar, onde permanece constante, conforme à Figura 6.

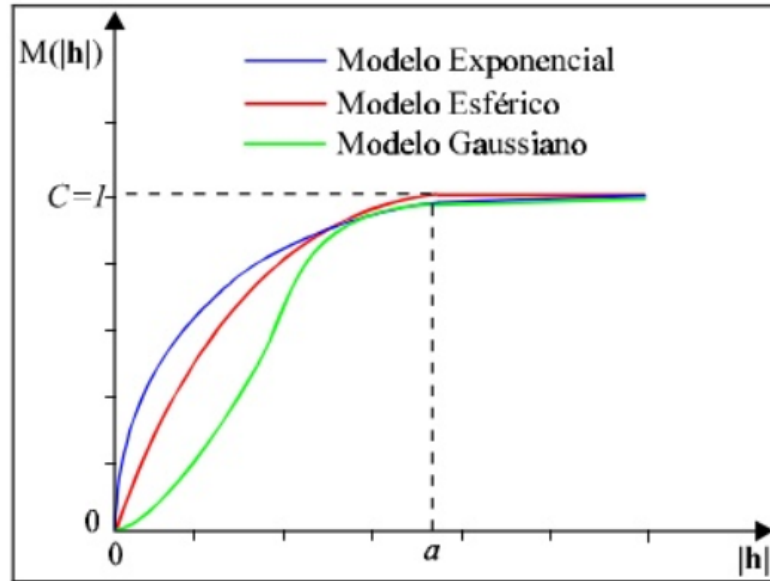


Figura 6: Representação gráfica de modelos com patamar (Monteiro et al., 2004a).

- **Modelo Esférico ou Matheron:** Neste ajuste do modelo ao variograma experimental temos que a tangente pela origem atinge o valor do patamar aproximadamente $2/3$ da amplitude variográfica. Matematicamente sua fórmula é dada por:

$$\gamma(h) = \begin{cases} C \left[\frac{3h}{2a} \right] - \left[\frac{h^3}{2a^3} \right] & , \text{ para } h \leq a \\ C & , \text{ para } h > a \end{cases} \quad (7)$$

em que C é o patamar, que corresponde à variância a priori finita e a é a amplitude variográfica.

- **Modelo Exponencial:** O ajuste se faz através da curva exponencial. O patamar é a assíntota e a amplitude variográfica corresponde ao encontro da tangente da curva na origem com o patamar:

$$\gamma(h) = \begin{cases} C \left(1 - e^{\left(-\frac{h}{a}\right)}\right) & , \text{ para } h \leq a \\ C & , \text{ para } h > a \end{cases} \quad (8)$$

a inclinação da tangente da curva na origem é igual a C/a .

- **Modelo Gaussiano:** A função é parabólica próxima à origem e também apresenta um ponto de inflexão. Este modelo apresenta amplitude variográfica a qual é definida na distância onde o valor do variograma alcança 95% do patamar; e o patamar é semelhante ao modelo exponencial. Sua função é:

$$\gamma(h) = \begin{cases} C \left(1 - e^{\left(-\frac{3h^2}{a^2}\right)}\right) & , \text{ para } h \leq a \\ C & , \text{ para } h > a \end{cases} \quad (9)$$

Nos modelos esféricos, exponencial e gaussiano o efeito pepita é, via de regra, somado ao modelo variográfico como uma constante (Ramírez, 2009).

- **Modelo aleatório (Efeito Pepita Puro):** Este modelo pode ser considerado como um caso particular de um modelo esférico de alcance infinitamente pequeno, mas do ponto de vista físico, existe uma diferença fundamental entre ambos os modelos (Ramírez, 2009). O modelo representa uma regionalização descontínua, onde a medida que aumenta a descontinuidade na origem do semi-variograma, mais aleatório é o fenômeno que originou a variável em análise e, no caso do modelo esférico descreve uma regionalização contínua. Sua expressão é:

$$\gamma(h) = C \quad (10)$$

para qualquer h .

2.4.2.2 Modelos sem Patamar

Estes modelos consideram somente a hipótese intrínseca. Nestes casos, a variância à priori é uma função de incremento direto em relação à área ou ao volume, onde é calculada. Os semivariogramas podem ser definidos, mas apresentam um aumento constante da variabilidade à medida que a distância cresce. Exemplo Figura 7.

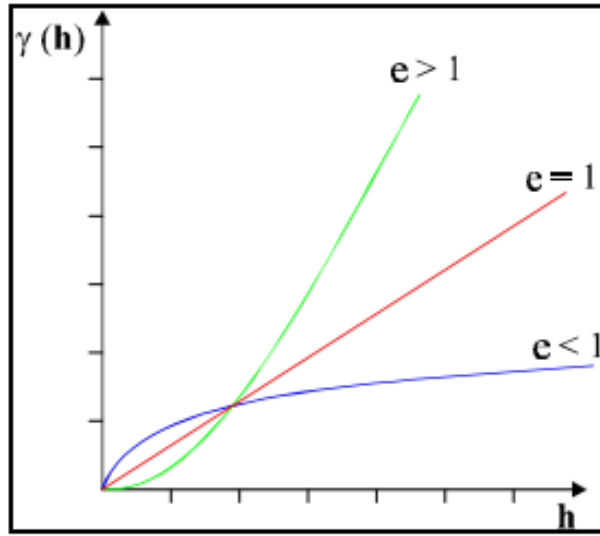


Figura 7: Representação gráfica de modelos sem patamar (Monteiro et al., 2004a).

- **Modelo Potência:** A expressão matemática geral para esses semivariogramas é:

$$\gamma(h) = \alpha h^e \quad (11)$$

em que e varia de 0 a 2 e α é a inclinação na origem.

- **Modelo Logarítmico (Esquema de Wigs):** Foi um modelo muito usado até 1966, quando surgiram os vários modelos com patamar. Sua equação, segundo a Lei de Wigs é:

$$\gamma(h) = 3\alpha \log_e(h) \quad (12)$$

em que α é uma constante que representa a dispersão absoluta.

2.5 Tipos de Krigagem

Há vários métodos de Krigagem propostos na literatura, sendo os principais, apresentados na sequência. A escolha do método mais apropriado depende do contexto dos dados na área de estudo. Nas seções 2.6 e 2.7 apresentamos, respectivamente, a Krigagem Ordinária e a Krigagem Indicativa, de grande importância no presente trabalho.

- **Krigagem Simples:** é utilizada quando a média é assumida como estatisticamente constante para toda a área.
- **Krigagem Ordinária:** utiliza médias locais ou tendências locais estimadas a partir dos elementos amostrais vizinhos, ao invés de uma única média estacionária. A média é constante, mas desconhecida.
- **Krigagem Indicativa:** requer uma transformação não linear, chamada de codificação por indicação, que transforma cada valor do conjunto amostral em valores por indicação, ou seja, a variável resultante da aplicação da função aleatória resultante não linear $f(z) = 0$ ou 1 .
- **Krigagem Universal:** é o método para os casos em que o processo estocástico não é estacionário, ou seja, o processo apresenta uma tendência.
- **Krigagem Disjuntiva:** assume o modelo com uma função arbitrária da função $Z(x)$, e com este método pode-se prever tanto o valor quanto o indicador, uma vez que a Krigagem de indicadores é um caso especial da Krigagem Disjuntiva.

2.6 Krigagem Ordinária

Esta metodologia é uma técnica de estimativa utilizada quando valores de uma variável regionalizada tem o valor médio constante, porém desconhecido. O algoritmo da Krigagem Ordinária é aplicado para encontrar os ponderadores ótimos que minimizem a variância dos erros de estimação.

Matematicamente, seja um ponto que se deseja estimar o valor da variável; seu valor real é desconhecido e representado por V , e seja S o conjunto dos dados amostrados localizados em coordenadas conhecidas. A estimativa de uma variável em local não amostrado (V^*) é calculado utilizando a combinação linear dos n dados, com valores $x_1, x_2, x_3, \dots, x_n$ (conjunto S), disponíveis na vizinhança do ponto analisado, de acordo com a expressão geral:

$$V^* = \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 + \dots + \lambda_n x_n \quad (13)$$

em que os λ_i são os pesos atribuídos a cada amostra i .

Existe associado a esse estimador um resíduo, $r = V^* - V$, isto é, a diferença entre o valor estimado e o valor verdadeiro (não conhecido); a melhor estimativa é obtida quando as condições de não tendenciosidade e de variância mínima são satisfeitas, conforme as expressões (15) e (16), respectivamente (Bagnara et al., 2012):

$$E(V^* - V) = 0 \quad (14)$$

$$Var(V^* - V) = \text{mínima} \quad (15)$$

A primeira condição implica a soma dos pesos for igual à unidade:

$$\sum_{i=1}^n \lambda_i = 1 \quad (16)$$

A minimização da variância do resíduo, sob as condições da soma dos pesos igual a 1, equação (17), e não ocorrer tendência local dos valores, conduz á dedução de um sistema de equações lineares com n incógnitas ($\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$), denominado equações da krigagem ordinária:

$$\sum_{j=1}^n \lambda_j \gamma_{ij} - \mu = \gamma_{i0}, \forall i = 1, 2, \dots, n \quad (17)$$

em que γ_{ij} e γ_{i0} são obtidos via variograma e correspondem, respectivamente, às semivariâncias calculadas entre pares de pontos amostrais e entre pontos amostrais e o ponto a ser estimado (Bagnara et al., 2012).

Como se tem apenas n incógnitas desconhecidas e $n + 1$ equações, introduz-se uma outra incógnita μ , também desconhecida, para balancear o sistema, conhecida multiplicador de Lagrange. A variância da estimativa pode ser calculada pela expressão:

$$\sigma_R^2 = \sum_{i=1}^n \lambda_i \gamma_{i0} + \mu \quad (18)$$

O objetivo da Krigagem é procurar pelo conjunto ótimo de ponderadores de modo que a variância do erro de estimativa seja a menor possível. Para tanto, é organizado um sistema de equações com $n + 1$ incógnitas, para a estimativa em um ponto (S_0) (Landim, 2006):

$$\lambda_1 \bar{\gamma}(S_1, S_1) + \lambda_2 \bar{\gamma}(S_1, S_2) + \lambda_3 \bar{\gamma}(S_1, S_3) + \dots + \mu = \bar{\gamma}(S_1, S_0)$$

$$\lambda_1 \bar{\gamma}(S_2, S_1) + \lambda_2 \bar{\gamma}(S_2, S_2) + \lambda_3 \bar{\gamma}(S_2, S_3) + \dots + \mu = \bar{\gamma}(S_2, S_0)$$

$$\vdots$$

$$\lambda_1 \bar{\gamma}(S_n, S_1) + \lambda_2 \bar{\gamma}(S_n, S_2) + \lambda_3 \bar{\gamma}(S_n, S_3) + \dots + \mu = \bar{\gamma}(S_n, S_0)$$

$$\lambda_1 + \lambda_2 + \lambda_3 + \dots + 0 = 1$$

Estas equações constituem equações normais a $n + 1$ incógnitas, as quais podem ser resolvidas, para a obtenção dos coeficientes, por cálculo matricial (Landim, 2006):

$$[S_i, S_j][\lambda_i] = [S_i, S_0] \quad (19)$$

Multiplicando ambos os termos da equação pelo inverso de $[S_i, S_j]$:

$$[S_i, S_j]^{-1}[S_i, S_j][\lambda_i] = [S_i, S_j]^{-1}[S_i, S_0] \quad (20)$$

temos pelas propriedades de matriz $[S_i, S_j]^{-1}[S_i, S_j] = [I]$ (matriz de identidade) e $[I][\lambda_i] = [\lambda_i]$,

$$[\lambda_i] = [S_i, S_j]^{-1}[S_i, S_0]. \quad (21)$$

Em notação matricial, temos que $[S_i, S_j]$ são os valores obtidos no semivariograma referentes às distâncias entre as amostras; o vetor $[S_i, S_0]$ são os

valores obtidos no semivariograma referentes às distâncias entre cada amostra e o ponto (S_0) a ser estimado e o vetor $[\lambda_i]$ contém os ponderadores a serem calculados:

$$\begin{array}{ccccc} \left[\begin{array}{ccccc} \bar{\gamma}(S_1, S_1) & \bar{\gamma}(S_1, S_2) & \dots & \bar{\gamma}(S_1, S_n) & 1 \\ \bar{\gamma}(S_2, S_1) & \bar{\gamma}(S_2, S_2) & \dots & \bar{\gamma}(S_2, S_n) & 1 \\ \vdots & \vdots & & \vdots & \\ \bar{\gamma}(S_n, S_1) & \bar{\gamma}(S_n, S_2) & \dots & \bar{\gamma}(S_n, S_n) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{array} \right] & \left[\begin{array}{c} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{array} \right] & = & \left[\begin{array}{c} \bar{\gamma}(S_1, S_0) \\ \bar{\gamma}(S_2, S_0) \\ \vdots \\ \bar{\gamma}(S_n, S_0) \\ 1 \end{array} \right] & (22) \\ [S_i, S_j] & [\lambda_i] & & [S_i, S_0] & \end{array}$$

Após resolver o sistema de equações, temos as estimativas dos pesos λ_i e o multiplicador de Lagrange, μ :

$$[\lambda_i] = [S_i, S_j]^{-1} [S_i, S_0] \quad (23)$$

Para estimar o ponto S_0 , obtém-se uma combinação linear dos valores dos pontos vizinhos e seus respectivos pesos:

$$S_0 = \sum_{i=1}^n \lambda_i S_i \quad (24)$$

Para o cálculo da variância (σ^2) associada ao valor S_0 obtido por estimativa usa-se a expressão:

$$\sigma^2 = \bar{\gamma}(S_i, S_0) + \mu = [\lambda_i]' [S_i, S_0] \quad (25)$$

onde $[\lambda_i]'$ é o vetor transposto com os pesos λ_i (Landim, 2006).

2.7 Krigagem Indicativa

A Krigagem Indicativa é um modelo não linear que consiste basicamente, na aplicação da Krigagem Ordinária para a variável transformada, que são variáveis contínuas transformadas em binárias, ou seja, ela modela a incerteza sobre o valor desconhecido $z^*(x)$ a partir da função de distribuição acumulada condicional, $F(x; I|(n_h))$, daquela variável.

Na Krigagem Ordinária básica, é feita uma estimativa para determinar valores médios, em locais não amostrados; na Krigagem Indicativa, as estimativas são estabelecidas através dos parâmetros de corte (*cutoff*) de uma distribuição acumulada, que irá indicar a probabilidade de um fenômeno ocorrer ou não, em um local estudado.

Este procedimento estabelecido aos dados transformados em indicadores para diversos níveis de corte (*cutoff*) de uma distribuição acumulada, conduz a função que poderá ser ajustada a probabilidade de uma variável resposta estar dentro ou fora dos parâmetros estabelecidos. Esta transformação não linear sobre o conjunto de dados amostrais é conhecida como codificação por indicação.

Inicialmente é preciso transformar os dados originais em indicadores, essa transformação indicativa é em valores binários, tendo apenas duas possibilidades 0 ou 1, pois os variogramas indicativos são mais simples de modelar e podem ser estimados pela função:

$$\gamma_i(h, v_c) = \frac{1}{2n_h} \sum_{i=1}^{n_h} [I(x_i, v_c) - I(x_i + h, v_c)]^2 \quad (26)$$

em que v_c é o nível de corte.

Na função (28), o caso de presença ou ausência, o valor 1 corresponde ao caráter presente e o valor 0 a ausência do caráter. Transformamos os valores que estão acima de um determinado nível de corte em zero (0) e os que estão abaixo, em um (1) (Fernandes, 2014):

$$I_j(v_c) = \begin{cases} 1 & , \text{ se } v_j \leq v_c \\ 0 & , \text{ se } v_j > v_c \end{cases} \quad (27)$$

em que v_j é o valor observado.

2.8 Análise da dependência espacial

Segundo Guerra & Salles (1998), quando dos estudos referentes à análise da continuidade espacial, uma consideração fundamental a ser feita é determinar o grau de aleatoriedade presente nos dados, o que pode ser obtido pela

fórmula $E = \frac{C_0}{C}$ (apud Fernandes, 2014). Uma regra usada para decidir sobre a aleatoriedade é a seguinte:

- $E < 0,15$: componente aleatória pequena;
- $0,15 \leq E \leq 0,30$: componente aleatória significativa;
- $E > 0,30$: componente aleatória muito significativa.

O extremo dessa situação é quando não ocorre a covariância entre as amostras, não podendo aplicar a Krigagem, neste caso, dizemos que é o modelo de efeito pepita pura.

Outra forma para quantificar o grau de dependência espacial é calculando o Índice de Dependência Espacial (IDE) como proposto por (Zimback, 1994):

$$IDE = \frac{C}{C_0 + C} \cdot 100\% \quad (28)$$

em que C é o patamar e C_0 o efeito pepita.

Segue a classificação quanto ao grau de dependência espacial da variável em estudo (Cambardella et al., 1994):

1. Variável Independente Espacialmente, se a relação entre a componente estrutural e patamar for igual a 0%, o variograma será com o efeito pepita puro ou $\frac{C}{C_0+C} = 0$.
2. Variável com Fraca Dependência Espacial, se a componente estrutural for menor ou igual a 25% do patamar $\left(\frac{C}{C_0+C} \leq 0,25\right)$.
3. Variável com Moderada Dependência Espacial, se a componente estrutural representar entre 25% e 75% do patamar $\left(0,25 < \frac{C}{C_0+C} < 0,75\right)$.
4. Variável com Forte Dependência Espacial, se a relação entre componente estrutural e patamar estiver entre 75% e 100% $\left(0,75 < \frac{C}{C_0+C} < 1,00\right)$.

2.9 Processos Pontuais

De acordo com Bailey & Gatrell (1995), o método dos processos pontuais é considerado o modo mais simples de representar dados espaciais, cuja observação de interesse é a própria localização espacial dos eventos na área de estudo, objetivando estudar a distribuição espacial e a relação dos pontos, observando se existe um padrão espacial, e suavizar a superfície, calculando intensidade para cada ponto da área através de interpolação, sem modificar as características e variabilidade dos dados.

A tipologia do padrão de pontos conforme à Figura 8, são as seguintes:

- **Padrão Aleatório:** a distribuição espacial não apresenta qualquer padrão visível, ou seja, não apresenta qualquer modelo de distribuição.
- **Padrão Aglomerado ou *Clusterizado*:** há um padrão de similaridade, ou seja, os eventos estão concentrados no mesmo espaço dentro da área maior; há um padrão de pontos que mostra uma aglomeração.
- **Padrão Regular:** a distribuição dos pontos seguem um determinado padrão ou espaçamento, ou seja, os pontos estão espalhados em todo o espaço, mas mantendo uma regularidade entre eles.

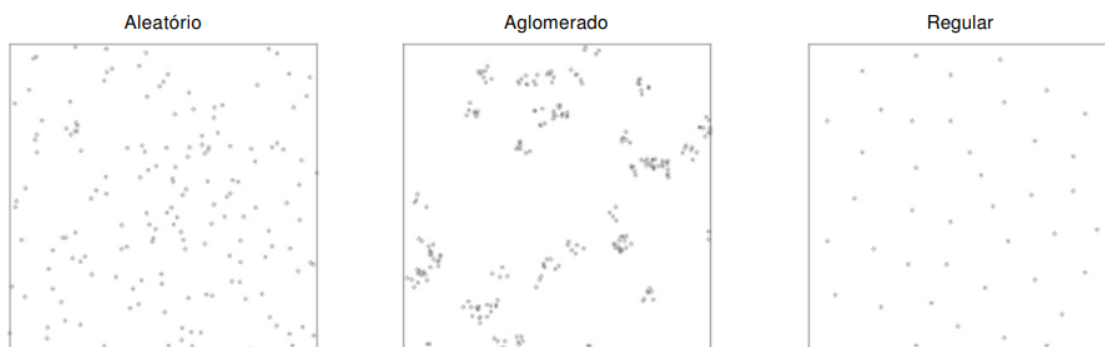


Figura 8: Ocorrência espacialmente aleatória, aglomerada e regular (Andrade et al., 2007).

Os processos pontuais são definidos como um conjunto de pontos distribuídos em uma área, cuja localização foi gerada por um mecanismo estocástico. Para sua caracterização, este processo estocástico pode ser descrito em termos dos efeitos de primeira ordem e efeitos de segunda ordem (Monteiro et al., 2004b):

- **Efeitos de primeira ordem:** considerados globais ou de larga escala, correspondem a variações no valor médio do processo no espaço. Neste caso, estamos interessados na intensidade do processo, isto é, no número de eventos por unidade de área em volta do evento.
- **Efeitos de segunda ordem:** denominados locais ou de pequena escala, representam a dependência espacial no processo, proveniente da estrutura de correlação espacial.

Quando estudamos processos pontuais, temos os eventos distribuídos no espaço, de modo que é possível analisar o comportamento de padrões de pontos e estimar a intensidade pontual do processo em toda a região de estudo. Dentre os procedimentos para analisar e estimar a intensidade dos pontos, a estimativa Kernel é a mais utilizada.

2.10 Estimador Kernel

É uma técnica de interpolação exploratória não paramétrica, que gera uma superfície de densidade para a identificação visual de “áreas quentes”, em que uma área quente é conhecida como a ocorrência de uma concentração de eventos que indica de alguma forma a aglomeração em uma região espacial (Bailey & Gatrell, 1995).

O estimador Kernel não é um método de detecção de aglomerados por si, mas, um método para explorar e mostrar o padrão de pontos de dados, e é útil a partir do momento em que gera uma superfície contínua a partir de dados pontuais (Craglia & Maheswaran, 2004), além disso, a técnica apresenta a vantagem

da rápida visualização de áreas que merecem atenção, além de não ser afetada por divisões político-administrativas.

A estimativa de Kernel foi originalmente desenvolvida para obter uma estimativa suave de uma densidade de probabilidade univariada ou multivariável de uma amostra observada, ou seja, um histograma suave (Bailey & Gatrell, 1995). O método é um interpolador, que concede a intensidade do evento em toda a área, mesmo nas regiões onde não há ocorrência real, ou seja, detectando a ocorrência de distribuições de eventos em aglomerados ou sua ocorrência de maneira aleatória; para isso, o Kernel é uma função de probabilidade bivariada adequadamente escolhida, na qual é simétrica em relação à origem (Bailey & Gatrell, 1995). Esta função realiza uma contagem de todos os pontos dentro de uma região de influência, ponderando-os pela distância de cada um à localização de interesse (Monteiro et al., 2004b).

Na Figura 9, vemos dois parâmetros básicos que são utilizados para a aplicação da estimativa Kernel e são definidos como (Andrade et al., 2007):

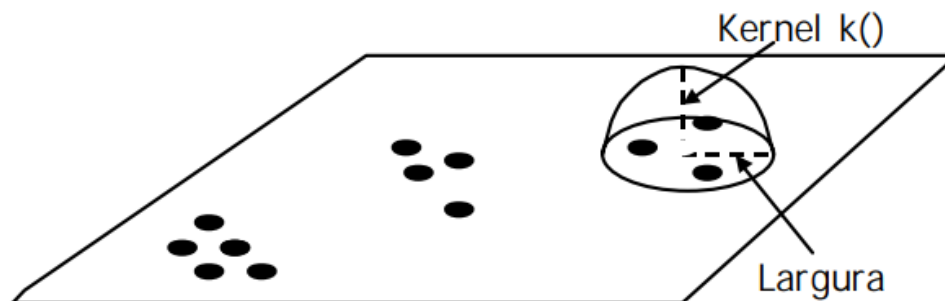


Figura 9: Estimador de intensidade de distribuição de pontos (Monteiro et al., 2004b).

- **Raio de Influência (σ):** Na Figura 9 descrito como largura, define a vizinhança do ponto a ser interpolado e controla o alisamento da superfície gerada. É o raio de um disco, centrado em s , que é uma localização na região R , no qual pontos s_i vão contribuir para a estimativa da função de intensidade. Um raio muito pequeno irá gerar uma superfície muito descontínua, e se for grande demais, a superfície poderá ficar muito suavizada (Matsumoto & Flores, 2012).

- **Função de estimação Kernel $k(\cdot)$:** é uma função matemática que tem propriedades de suavização do fenômeno. Os Kernel normal ou quártico são os mais comumente utilizados.

Se s representa uma localização em R e $s_1, \dots, s_n$ são as localizações das n observações, então a intensidade $\lambda(s)$, em s é estimada por (Bailey & Gatrell, 1995):

$$\hat{\lambda}_\sigma(s) = \sum_{i=1}^n \frac{1}{\sigma^2} k\left(\frac{(s - s_i)}{\sigma}\right) \quad (29)$$

Temos que k é uma função de densidade de probabilidade bivariada adequadamente escolhida. O parâmetro $\sigma > 0$ é conhecido como a largura de banda e determina a quantidade de suavização - essencialmente é o raio de um disco centrado em s dentro do qual os pontos s_i contribuirão significativamente para $\hat{\lambda}_\sigma(s)$, conforme Figura 10 (Andrade et al., 2007).

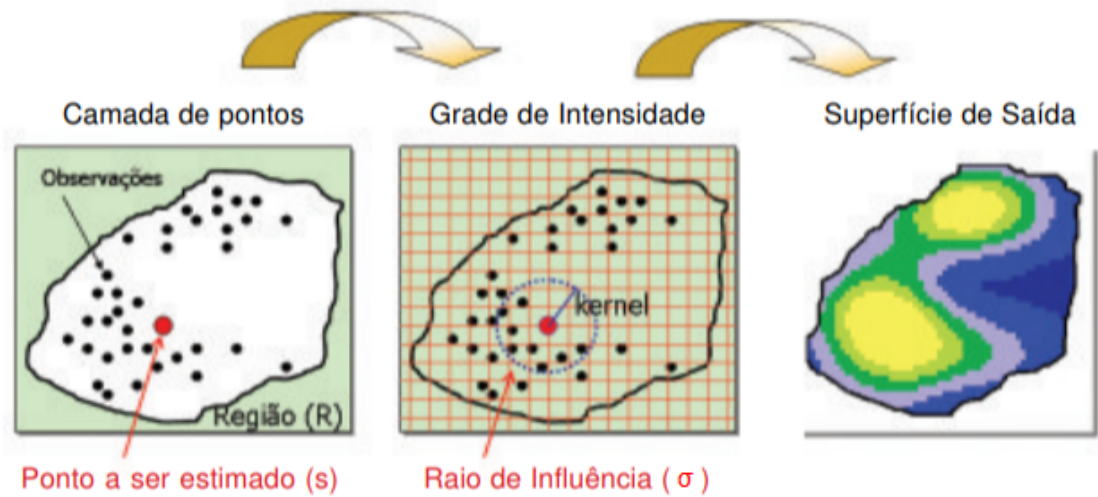


Figura 10: Passos para o cálculo de desenvolvimento de densidade de pontos segundo a técnica de Kernel (Andrade et al., 2007).

Quando os valores observados representam uma medida “média”, podemos ajustar a técnica de Kernel para estimar médias $\mu(s)$. Para toda posição do ponto cujo valor queremos estimar, o estimador de intensidade será computado a partir dos valores $y_1, \dots, y_n$ contidos no raio de tamanho σ . O Kernel com adição de

atributo é dado por:

$$\hat{\mu}_\sigma(s) = \frac{\sum_{i=1}^n k\left(\frac{(s-s_i)}{\sigma}\right) y_i}{\sum_{i=1}^n k\left(\frac{(s-s_i)}{\sigma}\right)} \quad (30)$$

A função $k()$ é um interpolador não paramétrico, por exemplo, o Kernel quártico é mostrado por:

$$k(h) = \begin{cases} \frac{3}{\pi}(1 - h^T h)^2 & , \text{ se } h^T h \leq 1 \\ 0 & , \text{ se caso contrário} \end{cases} \quad (31)$$

Observa-se que a estimativa do Kernel $\hat{\mu}_\sigma(s)$ é apenas uma versão mais sofisticada do esquema de média ponderada. Isto é, para um determinado σ , esta é essencialmente apenas uma média ponderada dos pontos de dados da amostra $\sum w_i(s)y_i$, onde os pesos:

$$w_i(s) = \frac{k\left(\frac{(s-s_i)}{\sigma}\right)}{\sum_{j=1}^n k\left(\frac{(s-s_j)}{\sigma}\right)} \quad (32)$$

dependem de σ , de s através da distância entre s e s_i e do local de intensidade dos pontos amostrados na vizinhança de s através do denominador. As amostras observadas fornecem a distância de s obtendo mais peso na região R onde pontos de amostra são mais escassos do que onde eles são densos (Bailey & Gatrell, 1995).

A escolha da função Kernel a ser utilizada não é um ponto crítico, contudo, a escolha do raio de influência é crucial, produzindo significantes alterações da estimativa final. Os principais estimadores Kernel, função de estimação $k(.)$ são:

- **Kernel Gaussiano ou Normal:** $k(h) = \frac{1}{2\pi} e^{\left(\frac{-h^2}{2\sigma^2}\right)}$
- **Kernel Quártico:** $k(h) = \frac{3}{\pi}(1 - h^2)^2$
- **Kernel Triangular:** $k(h) = (1 - |h|)$
- **Kernel Exponencial Negativo:** $k(h) = \frac{1}{2\pi} e^{-h}$
- **Kernel Uniforme:** $k(h) = \frac{1}{2}$

As diferenças entre as funções de Kernel são que a distribuição normal dá mais peso aos pontos mais próximos do círculo de forma mais intensamente,

comparados aos mais afastados; já a distribuição uniforme dá o mesmo peso à todos os pontos dentro do círculo; na função quártica os pontos próximos pesam mais do que os pontos distantes, mas o decrescimento é gradual; a função triangular pesa pontos próximos mais do que pontos distantes dentro do círculo, mas o decrescimento é mais rápido. A função exponencial negativa pesa muito mais os pontos mais próximos do que pontos distantes.

2.11 Método do Vizinho Mais Próximo

Como a estacionariedade de segunda ordem, representa a dependência espacial no processo, devido a estrutura de correlação, a técnica de análise espacial que estuda esses efeitos é o Método do Vizinho Mais Próximo, criada em 1966 por Cover e Hartque, na qual é muito eficiente para problemas de classificação de padrões pois consiste em estudar a distribuição espacial entre os eventos pontuais.

Uma abordagem simples para resumir o padrão usando distâncias mais próximas evento-evento, W , ou aquelas para distâncias ponto-evento, X , é estimar a distribuição cumulativa de probabilidade empírica ou função de distribuição, $\hat{G}(w)$ de W , ou $\hat{F}(x)$ de X dos dados como :

$$\hat{G}(w) = \frac{\#(w_i \leq w)}{n} \quad (33)$$

e

$$\hat{F}(x) = \frac{\#(x_i \leq x)}{m} \quad (34)$$

em que $\#$ significa ‘o número de’ e n é o número de eventos na área de estudo, ou m é o número de pontos aleatórios amostrados (Bailey & Gatrell, 1995).

O teste de significância utilizando o método vizinho mais próximo, pode ser avaliado através de um modelo teórico denominado Aleatoriedade Espacial Completa (“Complete Spatial Randomness – CSR”), que compara a distribuição dos eventos observados com o que se espera na hipótese CSR, criando envelopes de simulação para a distribuição CSR, afim de verificar a significância dos desvios.

Na hipótese de CSR, a função de distribuição $G(w)$ é um processo de Poisson, dado por (Bailey & Gatrell, 1995):

$$G(w) = 1 - e^{-\lambda\pi w^2}, w \geq 0 \quad (35)$$

A estimação simulada para a distribuição $G(w)$ assumindo-se CSR é calculada por (Bailey & Gatrell, 1995):

$$\bar{G}(w) = \sum_{i=1}^m \frac{\hat{G}_i(w)}{m}, \quad (36)$$

em que $\hat{G}_i(w)$, $i = 1, 2, \dots, m$ são funções de distribuição empíricas, estimadas a partir de m simulações independentes dos n eventos, na hipótese CSR (n eventos independentes e uniformemente distribuídos), o exemplo da Figura 11 mostra o comportamento do gráfico de $\hat{G}(w)$ conforme as distâncias dos pontos.

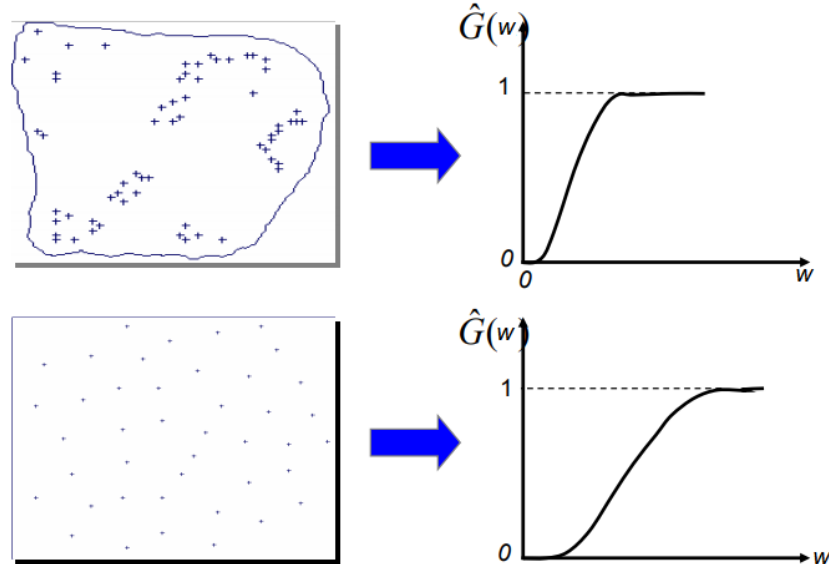


Figura 11: Gráfico de $\hat{G}(w)$ em relação as distâncias (Andrade et al., 2007).

Uma das opções é fazer o envelope de 95% de confiança, no qual o envelope superior é dado pelo percentil 97,5 e o envelope inferior é dado pelo percentil 2,5 das funções $G_i(w)$ simuladas, gerando o gráfico de tendência. Na Figura 12, podemos ver o comportamento das curvas no gráfico.

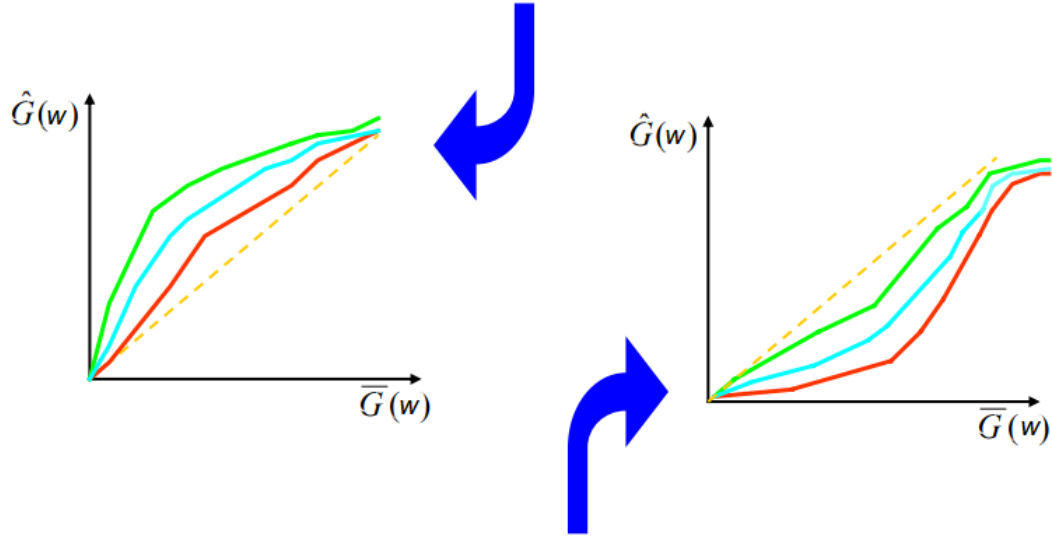


Figura 12: O gráfico à esquerda representa tendências para agrupamento e o gráfico à direita representa padrões de regularidade. Os traçados no gráfico estarão acima ou abaixo da linha de 45°, sendo a linha verde o envelope inferior, a linha azul o estimado e a linha vermelha o envelope superior (Vasconcelos, 2016).

Segundo Bailey & Gatrell (1995), um dos problemas em relação aos métodos de distância dos vizinhos mais próximos é que eles usam distâncias apenas para os eventos mais próximos, e portanto, consideram apenas o menor padrão de escala e as informações sobre escalas maiores são ignoradas.

Uma abordagem alternativa é usar a função K de Ripley que é um método de análise de segunda ordem baseado em contagem e distância, exigindo o conhecimento das coordenadas de cada ponto. Sua principal vantagem é que ele permite a detecção do padrão espacial em diferentes escalas de distâncias simultaneamente. Esta função permite ainda que sejam feitos testes de hipóteses sobre o padrão, estimar parâmetros e ajustar modelos.

A função K de Ripley é definida da seguinte maneira:

$$K(h) = \lambda^{-1} E(\#(\text{eventos dentro da distância } h \text{ de um evento arbitrário})) \quad (37)$$

em que novamente $\#$ significa ‘o número de’, $E()$ é o seu operador de esperança e λ é a intensidade ou o número médio de eventos por unidade de área, assumindo

constante ao longo de R .

O valor prático de $K(h)$ como medida resumida de efeitos de segunda ordem, é fundamentado para a obtenção de uma estimativa direta; um estimador adequado é dado por $\hat{K}(h)$, de um padrão pontual observado, ao passo que não seria possível estimar diretamente a intensidade da segunda ordem:

$$\hat{K}(h) = \frac{A}{n^2} \sum_{i=1}^m \sum_{j=1}^m \frac{I_h(d_{ij})}{p_{ij}}, i \neq j \quad (38)$$

em que A é área de R , d_{ij} é a distância entre os eventos i -ésimo e j -ésimo observados, $I_h(d_{ij})$ é uma função indicadora, em que é 1 se $d_{ij} \leq h$ e 0 caso contrário, então o número observado de tais pares ordenados é $\sum_{i=1}^m \sum_{j=1}^m I_h(d_{ij}), i \neq j$, sendo h o raio da circunferência onde o círculo está centrado no evento i , passando pelo ponto j e p_{ij} é a probabilidade condicional de um evento observado em R , dado uma distância d_{ij} do i -ésimo evento.

O gráfico da função K não é intuitivo, portanto utiliza-se uma função auxiliar L , para facilitar a interpretação, cujo seu estimador é dados por:

$$\hat{L}(h) = \sqrt{\frac{\hat{K}(h)}{\pi}} - h \quad (39)$$

Na Figura 13, vemos a interpretação da plotagem de $\hat{L}(h)$, no qual é possível ver que tendência de padrão das amostras estão seguindo:

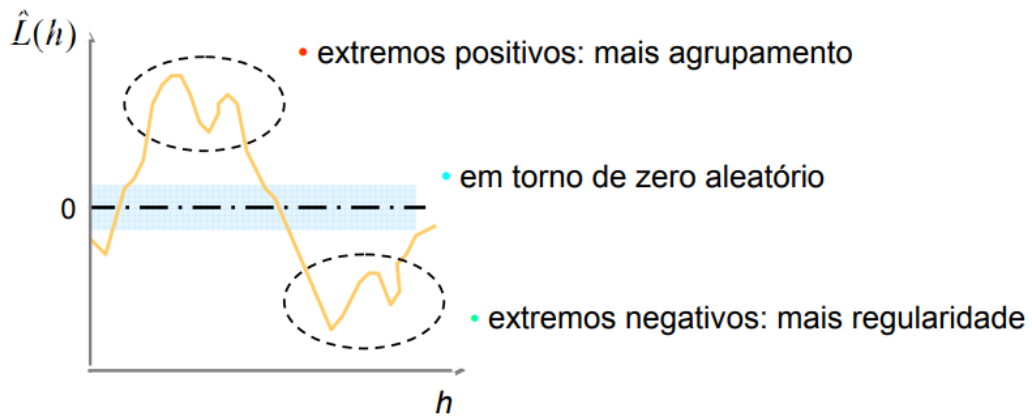


Figura 13: Gráfico e interpretação da plotagem de $\hat{L}(h)$ (Andrade et al., 2007).

Para o teste de significância dos desvios da distribuição $\hat{L}(h)$ em relação a aleatoriedade (CSR), as simulações são sobre a região e computam os envelopes superior e inferior dados por (Bailey & Gatrell, 1995), exemplo Figura 14:

$$Superior(h) = \max \{ \hat{L}_i(h) \}, i = 1, \dots, m \quad (40)$$

$$Inferior(h) = \min \{ \hat{L}_i(h) \}, i = 1, \dots, m \quad (41)$$

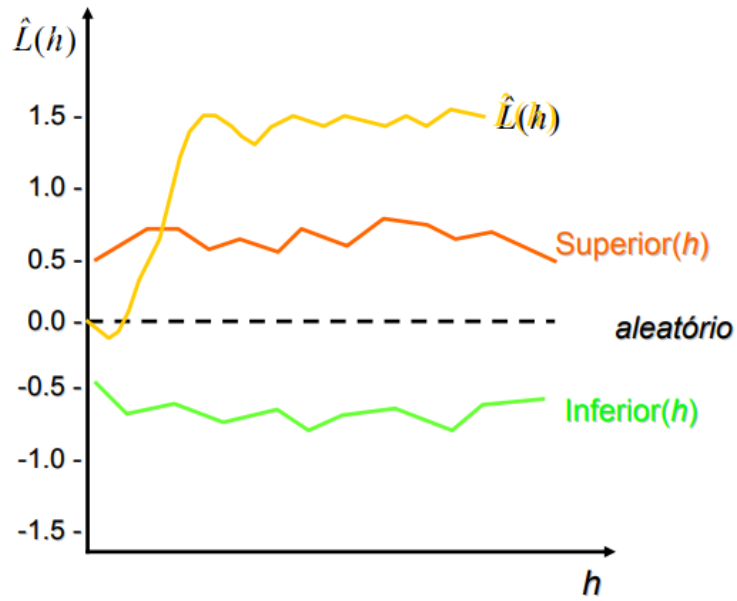


Figura 14: Análise do gráfico $\hat{L}(h)$ com os envelopes (Andrade et al., 2007).

Pode-se também assumir uma isotropia, isto é, uma unidade de distância na direção vertical tem o mesmo peso que uma unidade de distância horizontal. E como a função $k(h)$ fornece uma inferência em nível global sobre a área de estudo, esta medida também pode ser considerada em uma forma local para o i -ésimo evento (Martins, 2007).

3 MATERIAL E MÉTODOS

3.1 Material

Para o desenvolvimento deste trabalho, foram escolhidos dois bancos de dados. O primeiro refere-se às notificações de dengue confirmadas no ano de 2011 com suas respectivas coordenadas na área urbana de Rio Claro, São Paulo, Figura 15; o segundo refere-se às notificações de malária confirmadas no ano de 1999 com suas respectivas coordenadas na área urbana de Porto Velho, Rondônia, Figura 16. Em ambos bancos de dados, o objetivo é identificar locais com maior aglomeração de casos através das técnicas Kernel e Krigagem Indicativa.

3.1.1 Banco - Dengue

O município de Rio Claro que está situado na Depressão Periférica Paulista $22^{\circ}24'39''\text{S}$ $44^{\circ}33'39''\text{O}$ e possui uma área de 502 km^2 e altitude média de 612 m (Azevedo et al., 2011).

A área urbana da cidade possui uma topografia regular e plana em sua maioria, com altitudes que variam entre 550 m e 650 m. Com 186.253 mil habitantes em 2010, a cidade de Rio Claro está passando por um processo de fragmentação socioespacial, pois toda a sua estrutura urbana está sendo modificada drasticamente devido ao crescimento desordenado (Lourenço et al., 2001).

O município identificou a primeira presença do mosquito *Aedes aegypti* em 1996. Os vetores mais encontrados, no logradouro, são o *Aedes aegypti* e o *Aedes albopictus*, sendo este último, mais frequente nas áreas rurais e nos distritos municipais (CVE, 2010).

O banco de dados constitui de 1321 casos de dengue confirmados, distribuídos em 197 setores censitários, conforme à Figura 15, nos quais estão identificados através de coordenadas UTM.



Figura 15: Mapa da área urbana da cidade de Rio Claro-SP com os 197 setores censitários. Fonte: Aplicativo Google Earth Pro.

3.1.2 Banco - Malária

A capital rondoniense Porto Velho se localiza $8^{\circ}45'43''\text{S}$ $63^{\circ}54'14''\text{W}$ na parte oeste da Região Norte do Brasil, em uma área de $34090,962\text{ km}^2$ e uma área urbana de $41,7\text{ km}^2$, com 309.748 habitantes em 1999, onde é abrangida pela Amazônia Ocidental no Planalto Sul-Amazônico (Wikipédia, 2019).

Devido as condições ambientais inseridas no contexto amazônico, torna-a uma região endêmica. A capital Porto Velho é uma das cidades que houve casos de malária, identificados principalmente na área urbana (Simão, 2001).

Os dados foram retirados de Simão (2001), constitui de 3253 casos de

malária confirmados, distribuídos em 90 setores censitários, conforme à Figura 16, nos quais estão identificados através de coordenadas UTM.

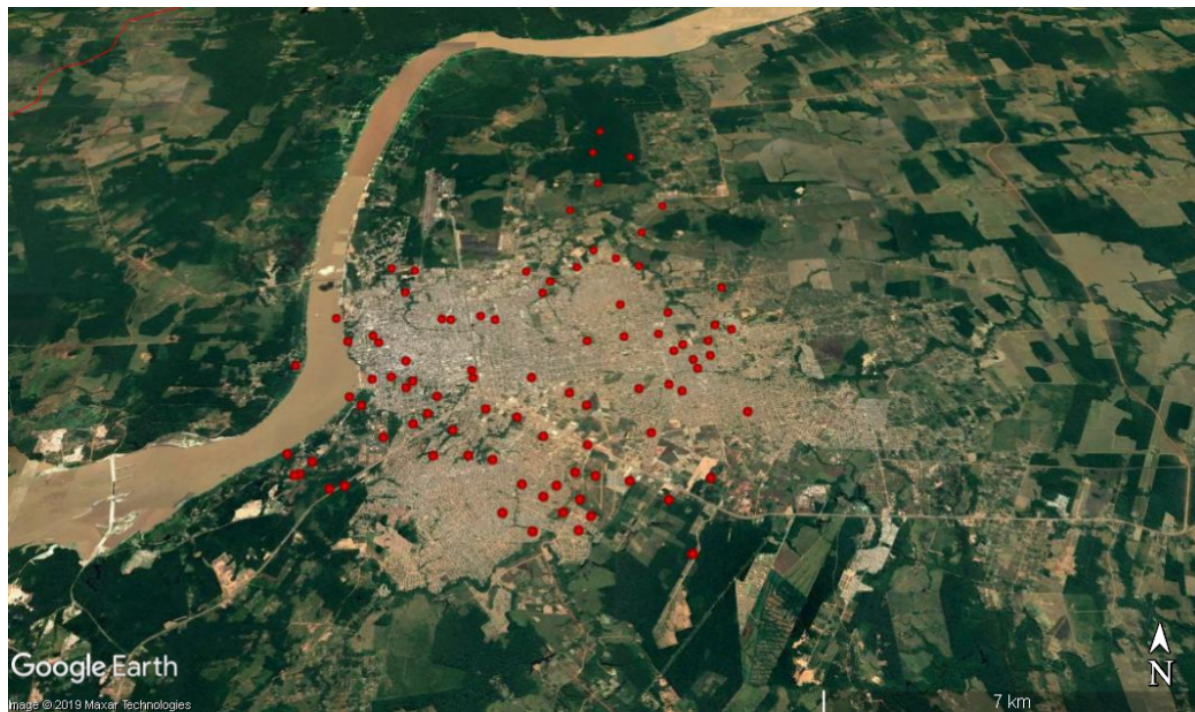


Figura 16: Mapa da área urbana da cidade de Porto Velho-RO com os 90 setores censitários. Fonte: Aplicativo Google Earth Pro.

Para aplicar as técnicas de Kernel e Krigagem Indicativa e fazer os testes foi utilizado o software gratuito R Core Team (2018).

3.2 Métodos

Os objetivos deste trabalho foram aplicar as técnicas Kernel e Krigagem Indicativa, em particularmente no campo dos estudos epidemiológicos, a fim de discutir as vantagens, desvantagens e compará-las para verificar qual teve melhor precisão nas previsões.

A escolha das técnicas se deu ao fato de ambas analisarem geograficamente o comportamento de padrões, utilizarem a autocorrelação espacial e gerar dados quanto a incerteza da previsão.

Inicialmente, para uma visão do comportamento dos atributos em estudo, fizemos as análises exploratória e exploratória espacial. Além disso, a detecção de *outliers* foi através do gráfico de *boxplot*. Para ambas análises utilizamos as funções de base e gráficos do software R Core Team (2018).

Em ambas as técnicas e em ambos os dados utilizando o pacote *stats* fizemos os critérios de informações Akaike (AIC) e Bayesiano (BIC).

Os critérios de informações são definidos como:

$$AIC = -2\ln L + 2p \quad \text{e} \quad BIC = -2\ln L + p \cdot \ln N \quad (42)$$

em que, $\ln L$ é a função de máxima verossimilhança, p é o número de parâmetros estimados e N número total de observações.

3.2.1 Krigagem Indicativa para os dados de dengue

Para aplicar a Krigagem Indicativa, é necessário escolher um ponto de corte para transformar os dados contínuos em binários, e para transformar os dados de dengue foi determinado um ponto de corte como a média das proporções dos casos em relação ao total, sendo 1 se for maior ou igual que o ponto de corte, e 0 se for menor, esse ponto foi escolhido porque pontos como a mediana não deram ajuste ao semivariograma.

Como o conjunto binário é um modelo estacionário de segunda ordem, fizemos o semivariograma utilizando o pacote *geoR* (Ribeiro Jr & Diggle, 2018), no qual, o testamos para a distância máxima, a metade da distância máxima, 1/3 da distância máxima e 1/4 da distância máxima.

Pela análise gráfica optamos pelo semivariograma com 1/3 da distância máxima, pois apresentou um bom alcance e um bom efeito pepita, além de ter menor variação entre os pontos.

Escolhido o melhor semivariograma, realizamos a análise anisotrópica que mostrou as direções do fenômeno espacial.

Para um modelo variográfico que melhor ajusta ao comportamento

espacial, procura-se ajustar um modelo matemático que mais se aproxima da configuração do semivariograma. Para este conjunto, foi utilizado a função *eyefit* para fazer um ajuste inicial, no qual o modelo escolhido foi o exponencial, e em seguida melhoramos este ajuste inicial do semivariograma por Mínimos Quadrados Ordinários (OLS) e por Mínimos Quadrados Ponderados (WLS). Também fizemos ajustes por Máxima Verossimilhança (ML) e por Máxima Verossimilhança Restrita (REML), neste caso não foi necessário ajustar um modelo inicial, pois estes já utilizam o modelo Gaussiano. E para saber qual foi o melhor, além da visualização gráfica, fizemos a autovalidação dos ajustes e comparamos as tabelas.

Escolhido o ajuste, obtemos a modelagem do semivariograma e aplicamos a Krigagem. Esta função executou a previsão espacial para parâmetros de covariância fixos usando vizinhança global.

Para entender melhor o comportamento dos dados, foram analisados diversos mapas. Primeiro elaboramos o mapa de isolinhas, que conecta os pontos em que a função tem um mesmo valor constante. Depois, foram feitos os mapas de Krigagem Indicativa e para verificar se a Krigagem Indicativa se aplicou bem, fizemos o mapa da variância da Krigagem, no qual foi analisado a média, a variância, o coeficiente de variação e o erro padrão.

3.2.2 Kernel para os dados de dengue

Primeiramente pelo pacote *spatstat* (Baddeley et al., 2015), calculamos as estatísticas de ponto espacial pela função *envelope.ppp*, que calcula envelopes de simulação de uma função de resumo. Sendo a função K de Ripley, para saber se os dados são distribuídos aleatoriamente; e a função F, em que é a função de espaço vazio ou sua taxa de risco.

Em seguida, a aplicação da técnica kernel depende criticamente do raio de influência empregado, em que não seja muito pequeno, pois irá gerar uma superfície muito descontínua e se for grande demais, a superfície poderá ficar muito suavizada.

Para o conjunto de dados de dengue, fizemos a estimação por validação cruzada, obtendo o raio de influência para a estimação da densidade kernel. O raio é escolhido como o valor mais alto, a fim maximizar o critério de validação cruzada.

Escolhido o raio de influência, fizemos a densidade da estimação Kernel, usando a função *density.ppp*, este que é um método para a densidade da função, que utiliza como padrão a função de estimação Kernel Gaussiano com o raio de influência encontrado pela validação cruzada.

Após a estimação da densidade Kernel, geramos o mapa de perspectiva da superfície e o mapa de incerteza (erro padrão).

3.2.3 Krigagem Indicativa para os dados de malária

Assim como no banco de dados anterior, transformamos os dados contínuos em binários. Seguindo as análises de (Simão, 2001) que para os casos de malária escolheu os pontos de corte como sendo os percentis, o ponto de corte escolhido para este trabalho foi o 25º percentil, pois foi o que obteve menor variância. Neste caso, os pontos menor ou igual ao ponto de corte foram transformados em 1 e, em 0 caso contrário.

Utilizando o pacote *geoR* (Ribeiro Jr & Diggle, 2018), fizemos o semi-variograma para diversas distâncias e, pela análise gráfica, o que apresentou melhor alcance, efeito pepita e menor variação entre os pontos foi a distância de 1/4 da distância máxima.

Fizemos a análise anisotrópica e em seguida procuramos um modelo matemático que se aproximasse da configuração do semivariograma escolhido. Utilizando a função *eyefit*, fizemos um ajuste inicial e o modelo escolhido foi o esférico, e após melhoramos o ajuste do semivariograma por Mínimos Quadrados Ordinários (OLS) e por Mínimos Quadrados Ponderados (WLS). Também fizemos os ajustes por Máxima Verossimilhança (ML) e por Máxima Verossimilhança Restrita (REML), porém neste caso não foi necessário fazer um ajuste inicial, pois é utilizado o modelo Gaussiano.

Escolhemos o ajuste através da visualização gráfica e da autovalidação. Após a modelagem do semivariograma, aplicamos a Krigagem e construímos mapas nos quais são possíveis ver as probabilidades acumuladas.

Primeiro elaboramos o mapa de isolinhas, depois foram feitos os mapas de Krigagem Indicativa e para verificar se a Krigagem Indicativa se aplicou bem, o mapa da variância da Krigagem, no qual, assim como anteriormente foi analisado a média, a variância, o coeficiente de variação e o erro padrão.

3.2.4 Kernel para os dados de malária

Assim como os dados de dengue, para os dados de malária, utilizamos o pacote *spatstat* (Baddeley et al., 2015), no qual calculamos as estatísticas de ponto espacial pela função *envelope.ppp*, gerando as funções K de Ripley, e a função F referente ao espaço vazio ou sua taxa de risco.

Utilizamos a função *bw.ppl* que usa a validação cruzada de probabilidade para selecionar um raio de influência de suavização para a estimativa da densidade Kernel.

Encontrado o raio de influência, fizemos a densidade da estimação Kernel, e geramos o mapa de perspectiva da superfície e o mapa de incerteza (erro padrão) da densidade Kernel.

4 RESULTADOS

4.1 Análise exploratória dos dados

Como análise estatística preliminar fizemos as análises exploratórias das contagens de dados, calculando as principais medidas em relação às ocorrências dos casos de dengue e malária respectivamente, a fim de facilitar o entendimento do comportamento dos mesmos. Os resultados são apresentados na Tabela 1.

Tabela 1: Resultados das análises exploratórias, por setores, das contagens de casos de dengue confirmados no ano de 2011 em Rio Claro-SP e das contagens de casos de malária no ano de 1999 em Porto Velho-RO.

Dengue		Malária	
Número de amostras	1321	Número de amostras	3253
Número de set. censitários	197	Número de set. censitários	90
Distância mínima	83,00602	Distância mínima	114,8434
Distância máxima	11039,32	Distância máxima	14514,97
Média	6,7056	Média	34,4111
Mediana	3	Mediana	15
Variância	83,1578	Variância	1768,02
Coefficiente de variação (%)	135,99	Coefficiente de variação (%)	122,19
Curtose	10,78	Curtose	4,48
Assimetria	2,60	Assimetria	1,52

Na Tabela 1, observamos que para os casos de dengue e malária há uma assimetria positiva e um coeficiente de variação alto. Podemos verificar essa

variação na Figura 17. A distribuição de frequências para malária (Figura 17, à esquerda), o número de linhas de classes se deu através da raiz quadrada do número de amostras; e a distribuição de frequências para dengue (Figura 17, à direita) se deu através da regra de Sturges, obtendo os histogramas:

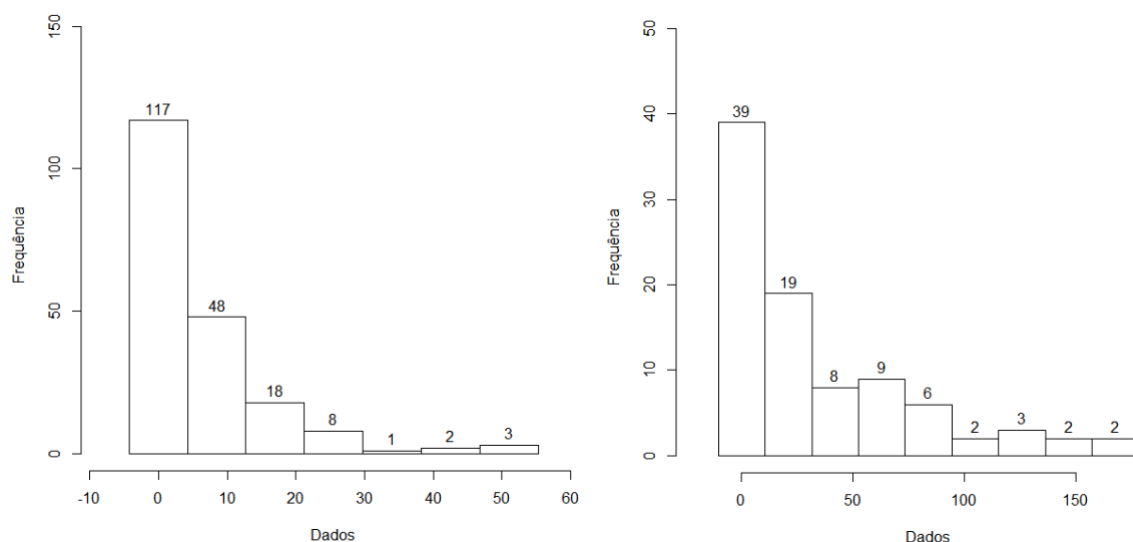


Figura 17: Histograma de frequência: à direita, da contagem de casos de dengue confirmados no ano de 2011 em Rio Claro-SP e à esquerda, da contagem de casos de malária confirmados no ano de 1999 em Porto Velho-RO.

Observa-se, em ambos histogramas, que as distribuições apresentaram grande variabilidade e assimetrias à direita, mostrando que as ocorrências em ambos os casos as distribuições dos dados não seguem uma normalidade, confirmando os dados da Tabela 1.

4.2 Análise exploratória espacial dos dados

4.2.1 Análise exploratória espacial dos dados de dengue

Na Figura 18, verificamos a distribuição espacial dos setores censitários da cidade de Rio Claro-SP, no qual não observamos qualquer padrão.

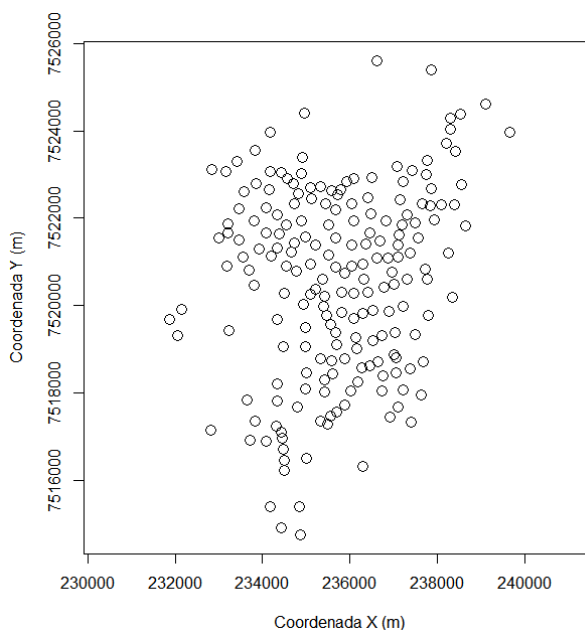


Figura 18: Distribuição espacial dos setores censitários da cidade de Rio Claro-SP.

Na Figura 19, observamos o gráfico de quantis em que a distribuição espacial dos setores censitários da cidade de Rio Claro-SP estão identificados de acordo com o quantil ao qual pertencem. Os gráficos coordenada X e coordenada Y são gráficos de dispersão da contagem de casos, estes gráficos permitem detectar a existência (ou não) de tendência, e neste caso há tendência. O histograma é referente a densidade da contagem de casos.

Outra informação importante para que se tenha uma visão do comportamento do atributo em estudo, é a detecção de *outliers*, pois alguns algoritmos desconsideram a presença de *outliers* em dados espaciais, ou apenas consideram sua localização em relação aos demais, gerando uma inconsistência nos resultados, neste caso pode ser visto através do gráfico de *boxplot*, na Figura 20.

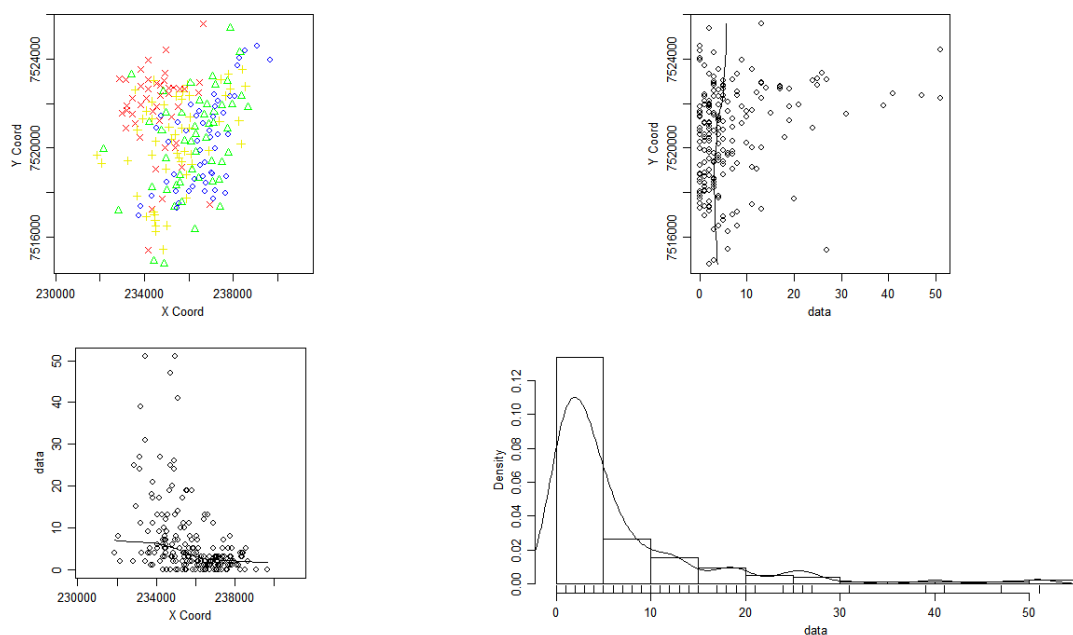


Figura 19: Análise de tendência casos de dengue confirmados no ano de 2011 em Rio Claro-SP. O primeiro gráfico de quantis está identificado de acordo com o quantil de cada setor censitário. O segundo e terceiro são gráficos de dispersão da contagem de casos e o quarto é o histograma referente à densidade da contagem de casos.

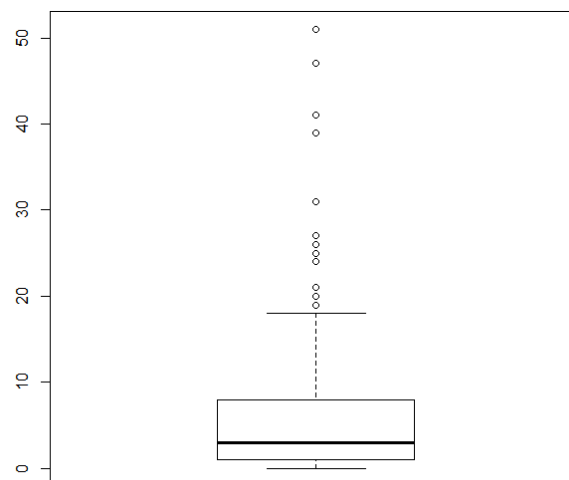


Figura 20: Gráfico boxplot da contagem de casos de dengue confirmados no ano de 2011 em Rio Claro-SP.

4.2.2 Análise exploratória espacial dos dados de malária

A Figura 21 mostra a distribuição espacial dos setores censitários da cidade de Porto Velho-RO, no qual não observamos qualquer padrão, ou seja, os setores estão distribuídos aleatoriamente.

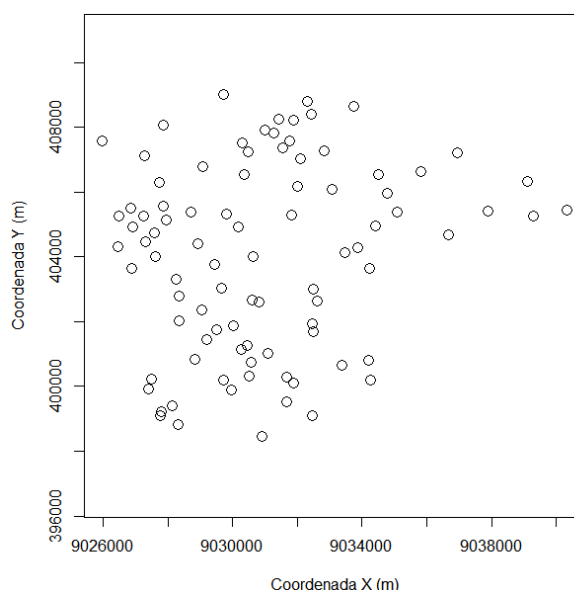


Figura 21: Distribuição espacial dos setores censitários da cidade de Porto Velho-RO.

Na Figura 22, observamos o comportamento dos dados de malária através do gráfico de quantis em que a distribuição espacial dos setores censitários da cidade de Porto Velho-RO, que estão identificados de acordo com o quantil ao qual pertencem. Os gráficos coordenada X e coordenada Y são gráficos de dispersão da contagem de casos, estes gráficos permitem detectar a existência (ou não) de tendência, e neste caso há tendência. O histograma é referente a densidade da contagem de casos.

O gráfico de *boxplot* da contagem dos casos de malária é visto na Figura 23, no qual verificamos que há poucos *outliers*.

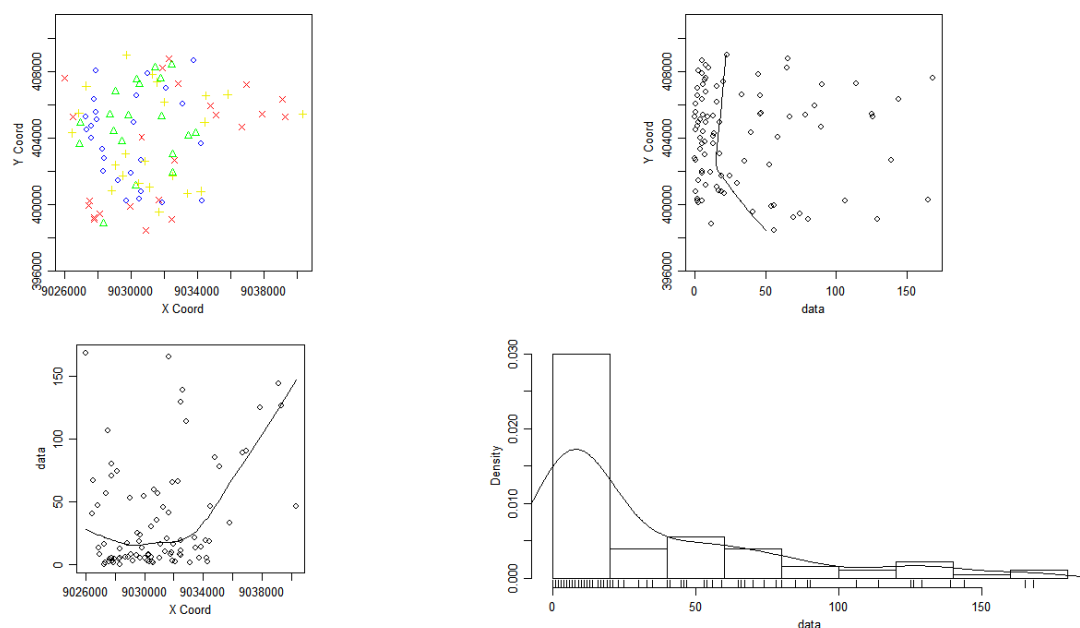


Figura 22: Análise de tendência casos de malária confirmados no ano de 1999 em Porto Velho-RO. O primeiro gráfico de quantis está identificado de acordo com o quantil de cada setor censitário. O segundo e terceiro são gráficos de dispersão da contagem de casos e o quarto é o histograma referente a densidade da contagem de casos.

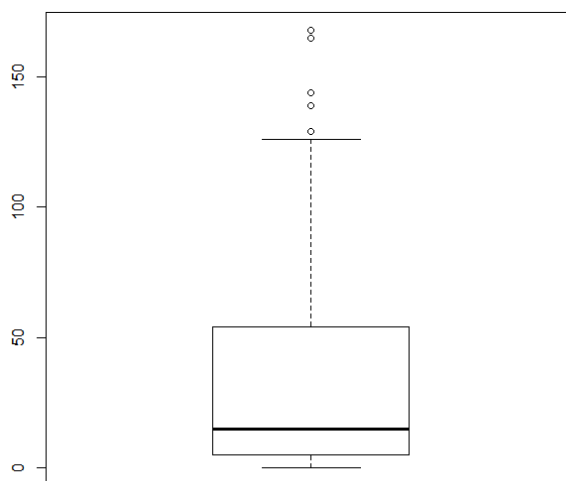


Figura 23: Gráfico boxplot da contagem de casos de malária confirmados no ano de 1999 em Porto Velho-RO.

4.3 Semivariograma dados de dengue

Com objetivo de descrever o comportamento da variabilidade espacial, para o processo de estimação, criamos semivariogramas indicativos. Para a transformação das variáveis aleatórias contínuas em binárias, para os dados de dengue, foi determinado um ponto de corte maior ou igual a média das proporções de casos em relação ao total, sendo 1 se o valor for maior que o ponto de corte, e 0 se for menor.

Observamos que os semivariogramas comportaram como um modelo estacionário de segunda ordem, e alternando a distância máxima obtivemos os resultados da Figura 24:

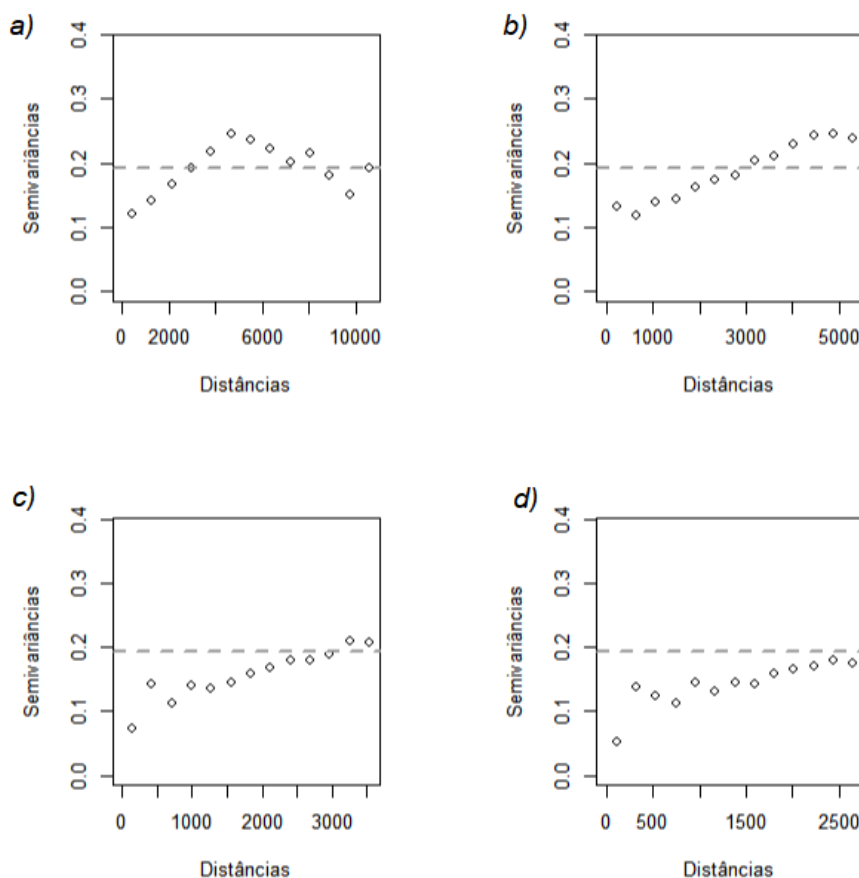


Figura 24: Semivariogramas dos dados binários de dengue, com as distâncias máximas: (a) distância máxima, (b) metade da distância máxima, (c) 1/3 da distância máxima e (d) 1/4 da distância máxima, respectivamente.

Pela análise gráfica da Figura 24, o melhor semivariograma foi com $1/3$ da distância máxima, pois apresentou o melhor alcance e um bom efeito pepita, além de ter menor variação entre os pontos.

Realizamos uma análise anisotrópica que mostrou as direções do fenômeno espacial, na Figura 25, em que observamos que os maiores valores de alcance estão nas direções de 0° e 45° e valores de maiores de patamar nas direções de 90° e 135° , sugerindo a presença de anisotropia combinada, que considera direção de menor e maior continuidade espacial.

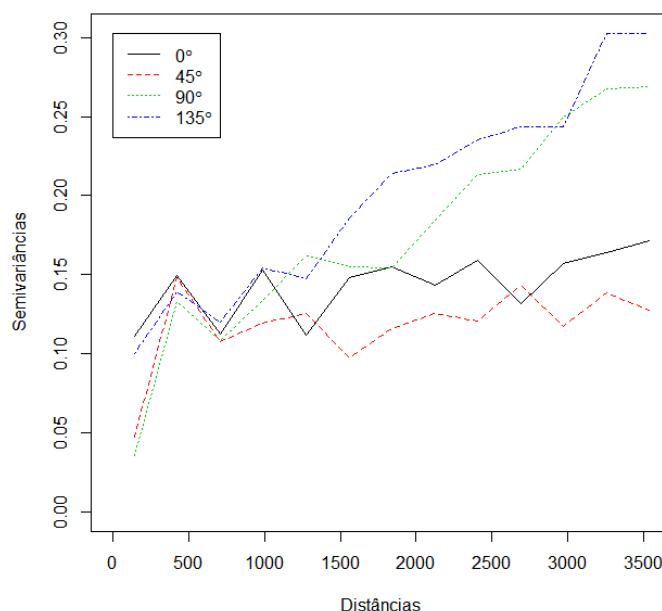


Figura 25: Gráfico da anisotropia dos dados binários dos casos de dengue.

Para um modelo variográfico que melhor se ajusta ao comportamento espacial, então procura-se ajustar um modelo matemático que mais se aproxima da configuração do semivariograma.

Usando a função *eyefit* do pacote *geoR* (Ribeiro Jr & Diggle (2018)), foi possível fazer um ajuste interativamente, por tentativa e erro. O modelo teórico escolhido foi o modelo exponencial e também foram escolhidos os parâmetros de ajuste: número de estruturas, efeito pepita, variância espacial e amplitude.

Com o ajuste inicial do semivariograma, o melhoramos através dos

Mínimos Quadrados Ordinários (OLS) e por Mínimos Quadrados Ponderados (WLS), e obtivemos os resultados da Figura 26:

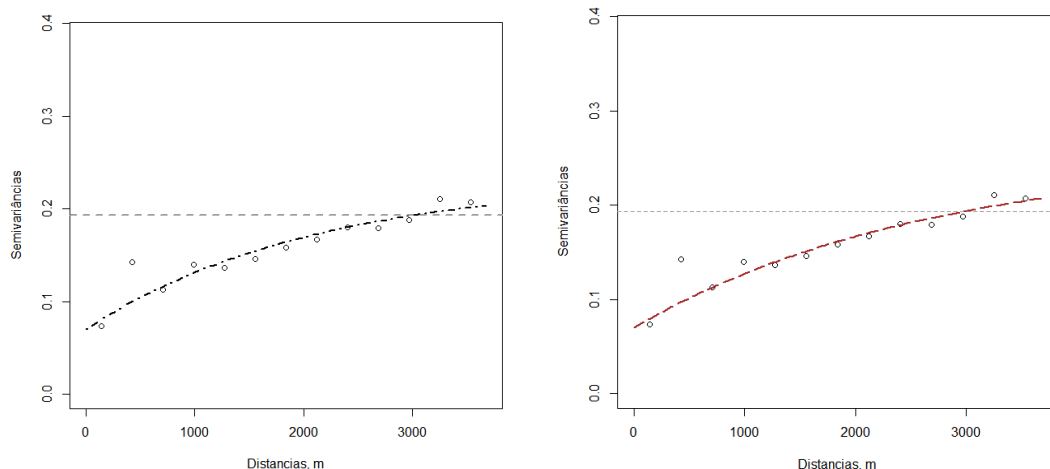


Figura 26: Gráficos dos ajustes dos semivariogramas: à esquerda com Mínimos Quadrados Ordinários (OLS) e à direita com Mínimos Quadrados Ponderados (WLS), para os casos de dengue confirmados no ano de 2011 em Rio Claro-SP.

Também foram feitos os ajustes por Máxima Verossimilhança (ML) e por Máxima Verossimilhança Restrita (REML) (Figura 27), neste caso não foi necessário fazer ajustes iniciais dos parâmetros, pois estes ajustes utilizam o modelo de ajuste Gaussiano.

Comparamos as Figuras 26 e 27, e os modelos ajustados graficamente ficaram muito parecidos, mas pelo mínimos quadrados se mostram um pouco melhor para o conjunto de dados. Fizemos a autovalidação de ambos os ajustes de mínimos quadrados, na Tabela 2, para comparação.

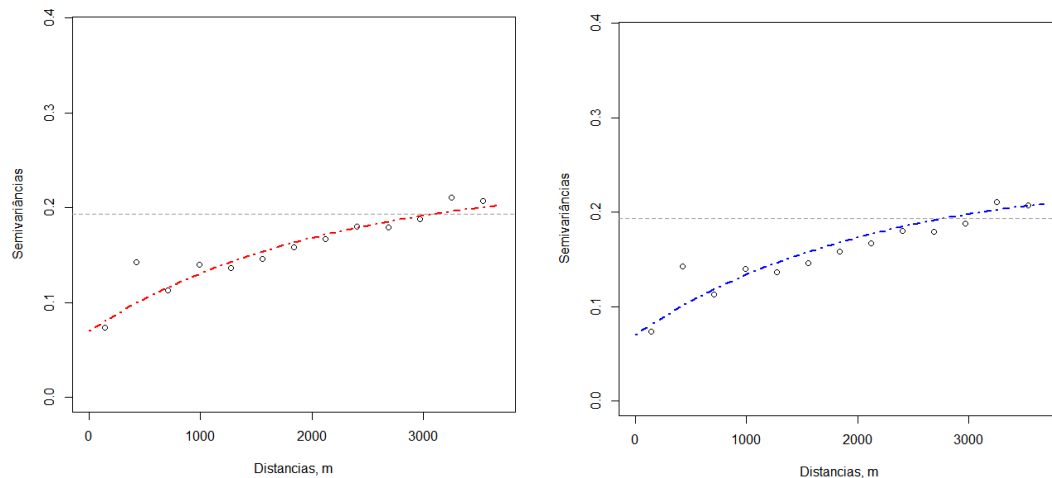


Figura 27: Gráficos dos ajustes dos semivariogramas: à esquerda com a Máxima Verossimilhança (ML) e à direita com a Máxima Verossimilhança Restrita (REML), para os casos de dengue confirmados no ano de 2011 em Rio Claro-SP.

Tabela 2: Autovalidação dos ajustes Mínimos Quadrados Ordinários (OLS) e Mínimos Quadrados Ponderados (WLS) para os dados de dengue.

	OLS	WLS
Média dos erros padronizados	0,00212	0,00196
Desvio dos erros padronizados	1,15866	1,17244
Variância dos erros	0,14645	0,14587
Média dos erros	-0,00149	-0,00138
β_0 da Regressão	0,04807	0,04761
β_1 da Regressão	0,81908	0,82048
R^2 ajustado	0,24909	0,25200
AIC	182,3772	181,6125
BIC	192,2268	191,4621

Em comparação na Tabela 2, os modelos se mostraram muitos parecidos, porém o modelo de Mínimos Quadrados Ponderados (WLS) se mostrou um pouco melhor como ajuste para o conjunto de dados de dengue, tendo valores me-

nores, como o AIC e BIC.

A Tabela 3 mostra os resultados obtidos pelo semivariograma, pelo método de Mínimos Quadrados Ponderados (WLS), e os valores de variâncias da Krigagem Indicativa para os casos de dengue, em que observamos que o modelo teve um efeito pepita baixo e um alcance prático alto, sendo um bom ajuste para este conjunto de dados.

Tabela 3: Resultados referentes a modelagem do semivariograma para à Krigagem Indicativa dos casos de dengue.

Modelagem do Semivariograma:	
Método de ajuste	Mínimos Quadrados Ponderados (WLS)
Alcance prático	8033,452
Contribuição	0,1837807
Efeito pepita	0,07
Modelo ajustado	exponencial

4.4 Krigagem Indicativa para os dados de dengue

Na Figura 28, temos o mapa de isolinhas que conectam os pontos em que a função tem um mesmo valor constante, em que cada valor de isolinha representa a probabilidade de haver casos de dengue.

No mapa da Krigagem Indicativa, Figura 29, vemos que a região mais clara mostra maior valor na região noroeste da cidade, isso representa os locais que tiveram maior probabilidade de ocorrência dos casos de dengue e os locais mais esverdeados foram os que tiveram a probabilidade baixa.

No mapa, Figura 30, observamos que houve baixa variância da Krigagem, em que a maior parte dos dados está com uma variância próximo à 0.02%, isso indica que houve pouca dispersão quando aplicada a técnica. Os dados da Tabela 4 também confirmam a baixa variância.

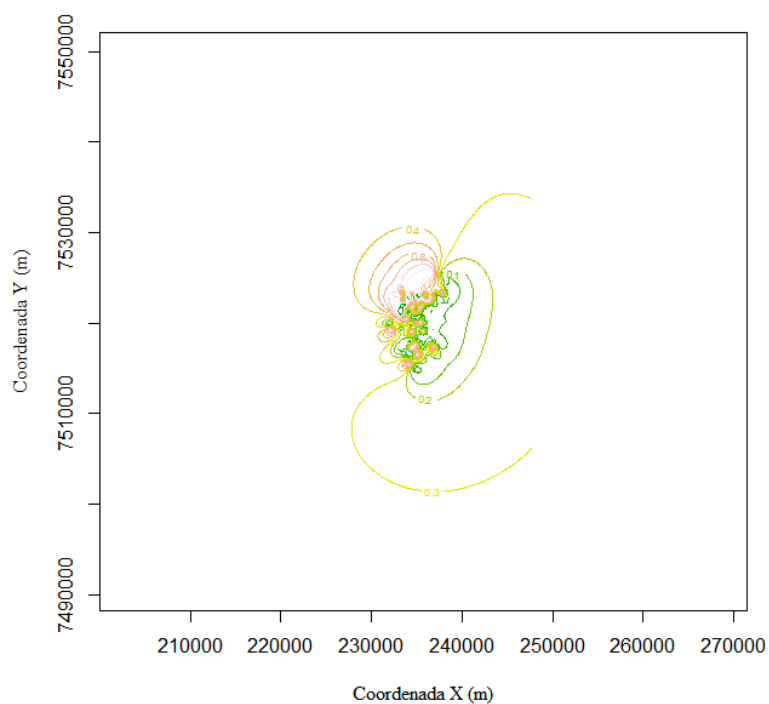


Figura 28: Mapa de isolinhas dos casos de dengue confirmados no ano de 2011 em Rio Claro-SP.

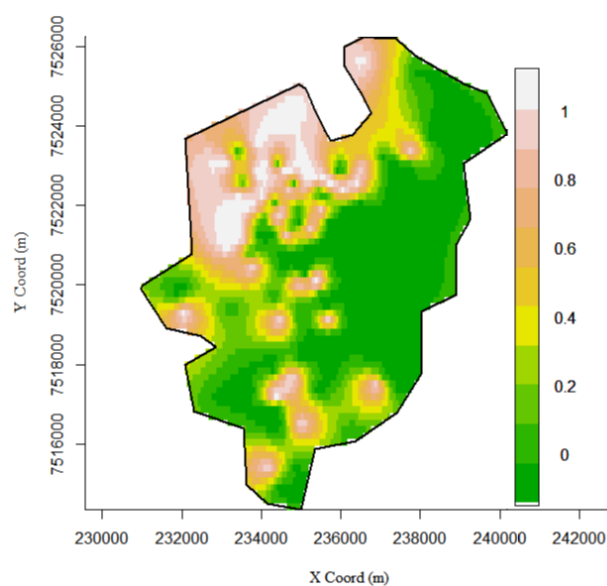


Figura 29: Mapa de Krigagem Indicativa referente aos casos de dengue confirmados no ano de 2011 em Rio Claro-SP.

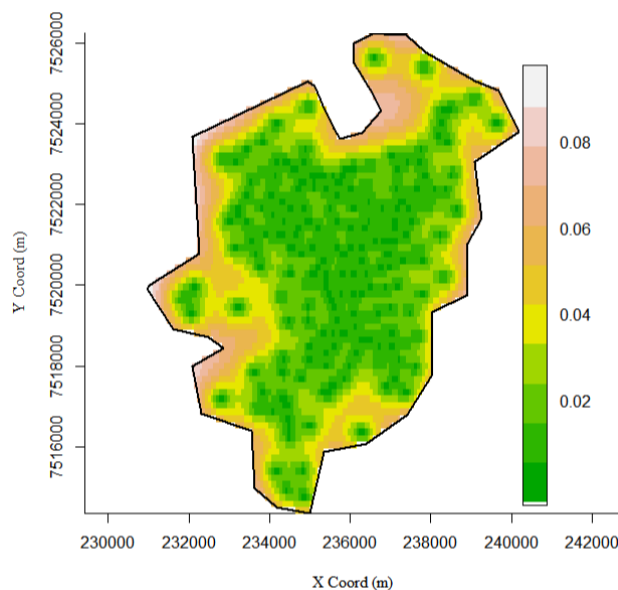


Figura 30: Variância de Krigagem referente aos casos de dengue confirmados no ano de 2011 em Rio Claro-SP.

Tabela 4: Variância de Krigagem Indicativa dos casos de dengue.

Média	0,0265431
Variância	0,0002601864
Coefficiente de variação	0,6077021
Erro padrão	0,001149236

4.5 Ajuste da função K de Ripley para os dados de dengue

No gráfico da função K de Ripley, temos a estimativa simulada pelo valor real da distância, Figura 31.

Sua principal vantagem é que ele permite a detecção do padrão espacial em diferentes escalas de distâncias simultâneas e aqui constatamos que os dados de dengue são concentrados.

O gráfico da função F, Figura 32, é a função de espaço vazio ou sua taxa de risco para os dados de dengue. Temos a estimativa simulada pelo valor real,

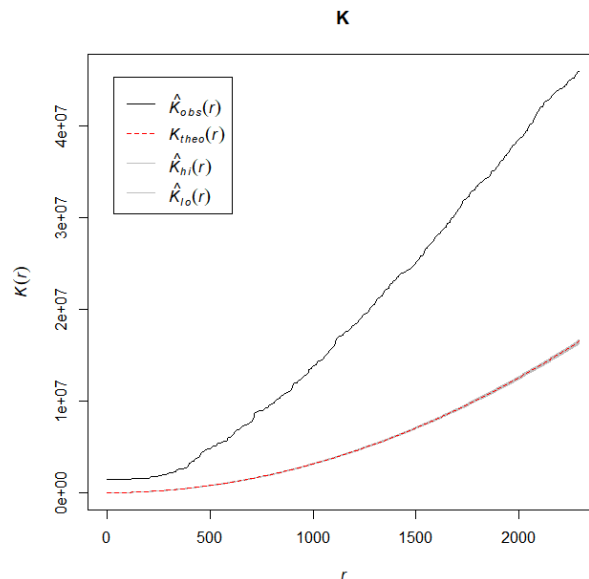


Figura 31: Gráfico da função K (função de Ripley) dos casos de dengue confirmados no ano de 2011 em Rio Claro-SP.

e observando o gráfico temos que o espaço vazio é baixo, mesmo quando há um aumento nas escalas.

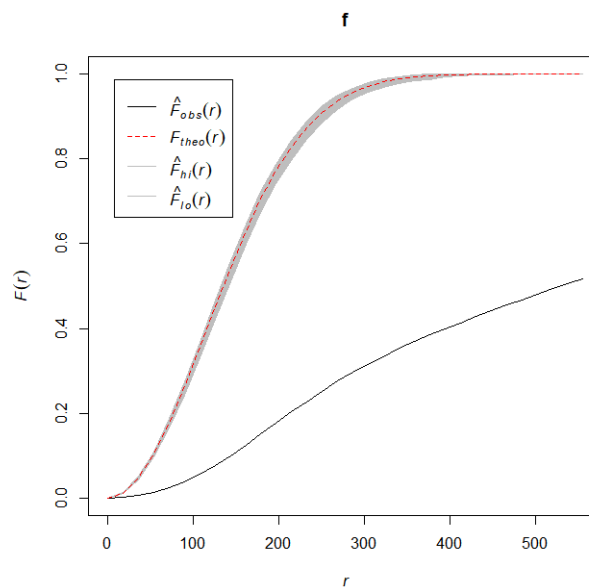


Figura 32: Gráfico da função F (função de espaço vazio ou taxa de risco) dos casos de dengue confirmados no ano de 2011 em Rio Claro-SP.

4.6 Kernel para os dados de dengue

Para a aplicação do Kernel precisamos primeiro achar o raio de influência, que seja adequado para a suavização. Um dos métodos de encontra-lo é pela validação cruzada.

O gráfico da Figura 33, mostra o raio de influência dado pela validação cruzada para os dados de dengue, em que o raio é escolhido como o valor mais, alto a fim maximizar o critério de validação cruzada e o raio de influência encontrado foi dado pelo valor de $\sigma = 241,0809$.

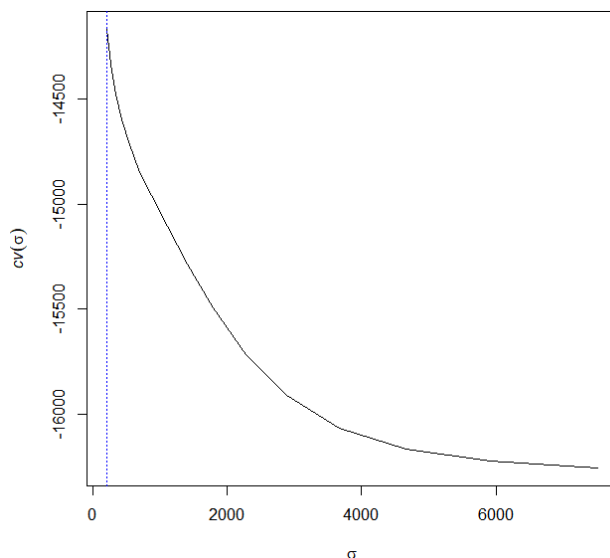


Figura 33: Gráfico do raio de influência, com $\sigma = 241,0809$, para os casos de dengue confirmados no ano de 2011 em Rio Claro-SP.

Escolhido o raio de influência, fizemos a densidade da estimacão Kernel, com o Kernel Gaussiano, gerando a seguinte superfície de densidade, Figura 34:

Podemos observar na Figura 34, que quanto mais clara a região, maior a concentração de casos e quanto mais escura, menor a ocorrência dos casos e neste caso a maior ocorrência de casos está na região noroeste.

Analizamos o mapa de incerteza, Figura 35, no qual quanto mais clara a região, maior a incerteza, porém algumas dessas incertezas associadas à estes dados estão no fato de que a resolução utilizada por esse sistema não permite distinguir as

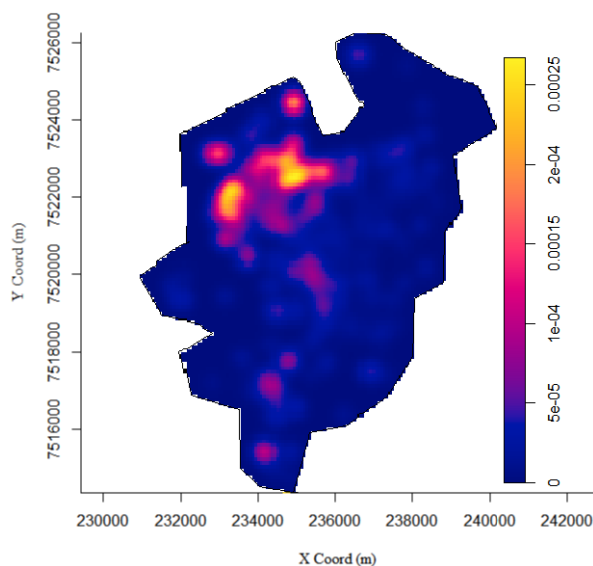


Figura 34: Mapa de estimaco da densidade Kernel referente aos casos de dengue confirmados no ano de 2011 em Rio Claro-SP.

reas onde ocorreu e ou no os casos, principalmente quando h uma concentrao dos dados, isso porque o mtodo interpolador concede a intensidade do evento em toda a rea mesmo nas regies onde no h ocorrncia real.

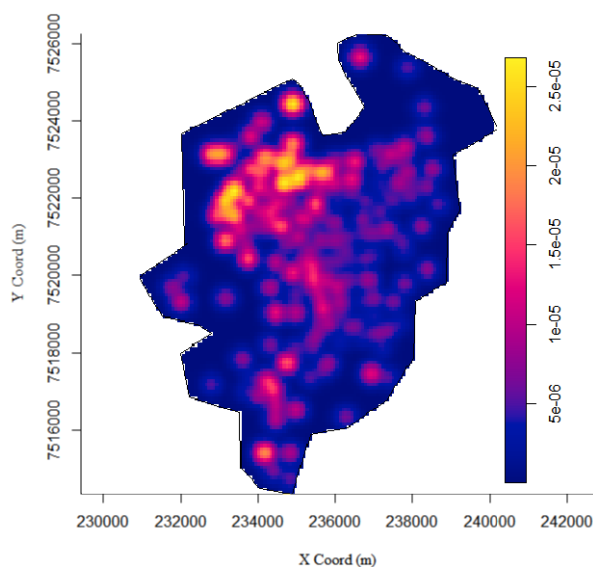


Figura 35: Mapa de incerteza da estimaco da densidade Kernel referente aos casos de dengue confirmados no ano de 2011 em Rio Claro-SP.

Para uma análise do desempenho da aplicação da técnica Kernel para este banco de dados, calculamos os critérios de informação Akaike (AIC) e Bayesiano (BIC), Tabela 5.

Tabela 5: Critérios de informação Akaike (AIC) e Bayesiano (BIC) da estimação de densidade Kernel dos casos de dengue.

AIC	-9,862476
BIC	-13.86248

4.7 Comparação das técnicas Kernel e Krigagem Indicativa para os dados de dengue

Com os resultados obtidos, observamos na Figura 36 que as técnicas Kernel e Krigagem Indicativa se aplicaram bem ao conjunto de dados de dengue, e foram eficientes na geração dos mapas, permitindo a fácil detecção das regiões onde houve concentração da endemia.

Comparando os mapas para os dados de dengue de ambas as técnicas, Figura 36, podemos ver que em ambos há maior ocorrência dos casos na região noroeste, porém houve maior concentração no mapa Kernel enquanto no mapa da Krigagem Indicativa mostra uma região maior dos casos. Visualmente, o mapa da técnica Kernel foi melhor, pois as regiões em que há focos se encontram mais definidas.

Lembrando que o mapa da Krigagem Indicativa representa a probabilidade do valor de corte a ser superado e o mapa kernel representa a intensidade de ocorrência em cada ponto.

Fizemos os critérios de informação e obtivemos os resultados da Tabela 6, e comparando-os, temos que o AIC e BIC para a técnica Kernel foram menores que a técnica Krigagem Indicativa, então concluímos que para os dados de dengue a técnica Kernel teve melhor precisão nas predições.

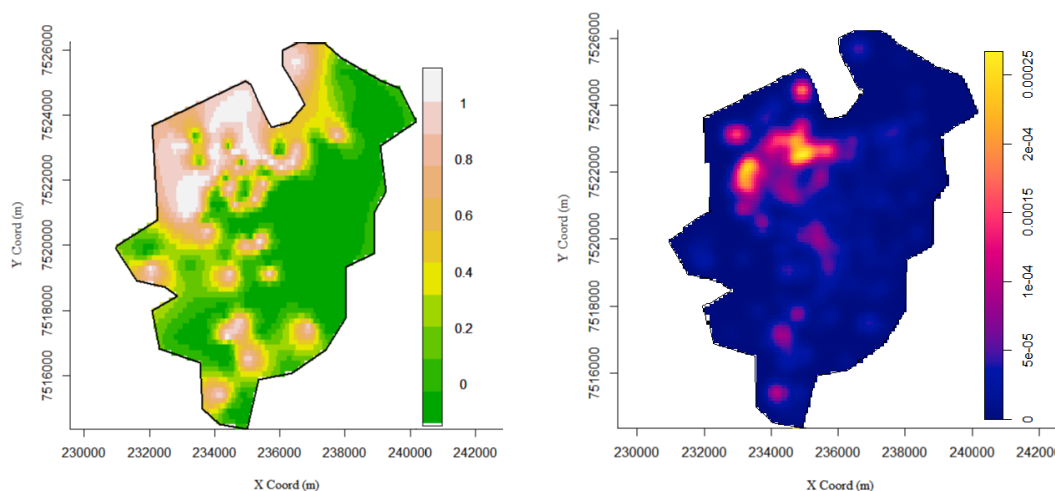


Figura 36: Comparação dos mapas para os dados de dengue, à direita o mapa Kernel e à esquerda o mapa da Krigagem Indicativa.

Tabela 6: Critérios de informação Akaike (AIC) e Bayesiano (BIC) da estimação de densidade Kernel e da Krigagem Indicativa dos casos de dengue.

	AIC	BIC
Krigagem Indicativa	181,6125	191,4621
Kernel	-9,862476	-13,86248

4.8 Semivariograma dados de malária

Na transformação das variáveis aleatórias contínuas em binárias para os dados de malária, foi determinado um ponto de corte menor ou igual ao 1º quartil (5 casos), sendo 1 se for menor ou igual ao ponto de corte, e 0 se for maior.

Os semivariogramas se comportaram como um modelo estacionário de segunda ordem. E alternando a distância máxima obtivemos os resultados da Figura 37:

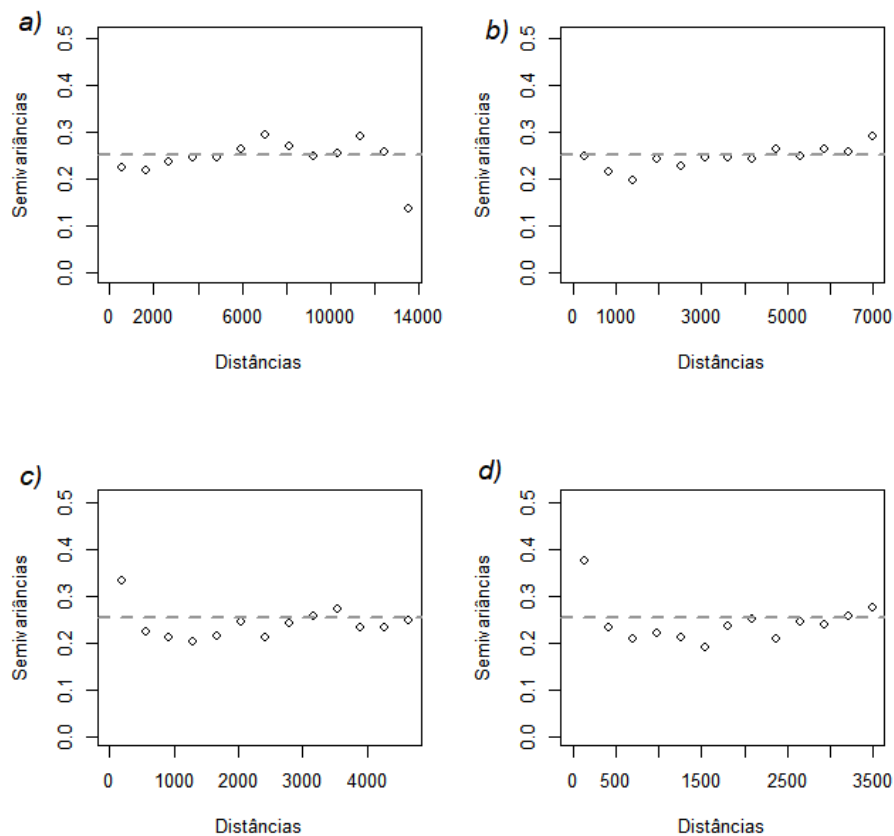


Figura 37: Semivariogramas dos dados binários de malária, com as: (a) distância máxima, (b) metade da distância máxima, (c) $1/3$ da distância máxima e (d) $1/4$ da distância máxima, respectivamente.

Pela análise gráfica da Figura 37, o melhor semivariograma foi com a distância máxima, que apresentou o melhor alcance e um bom efeito pepita, além de ter menor variação entre os pontos.

A análise anisotrópica mostrou as direções do fenômeno espacial, na Figura 38:

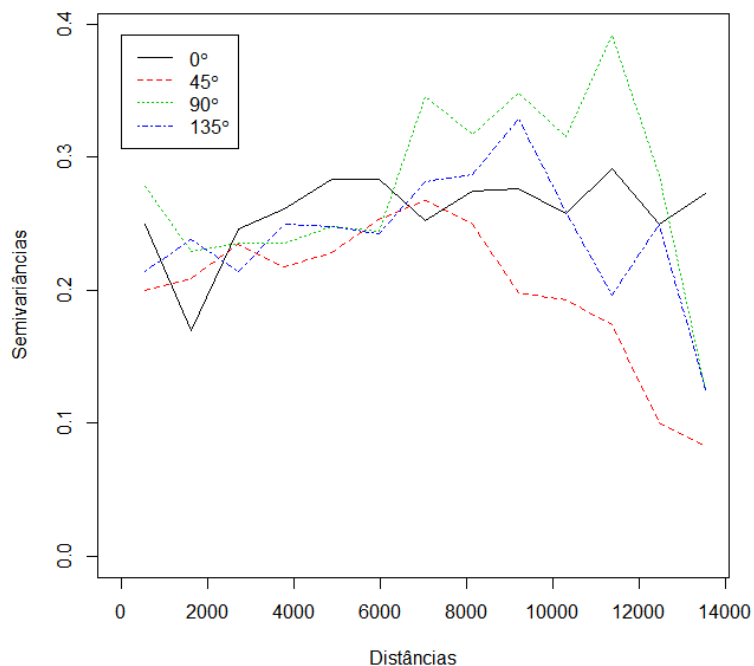


Figura 38: Gráfico da anisotropia dos dados binários dos casos de malária.

Observamos na Figura 38 que há uma isotropia, ou seja, a função semivariograma depende unicamente da magnitude do vetor, e não da direção.

Para um modelo variográfico inicial que melhor ajusta ao comportamento espacial, foi também utilizado a função *eyefit* do pacote *geoR* (Ribeiro Jr & Diggle (2018)), no qual foi escolhido os parâmetros de ajustes e o modelo esférico.

Com o ajuste preliminar, o melhoramos com os ajustes do semivariograma por Mínimos Quadrados Ordinários (OLS) e por Mínimos Quadrados Ponderados (WLS), Figura 39. A Figura 40 apresenta os ajustes por Máxima Verossimilhança (ML) e por Máxima Verossimilhança Restrita (REML) que utilizam o modelo Gaussiano.

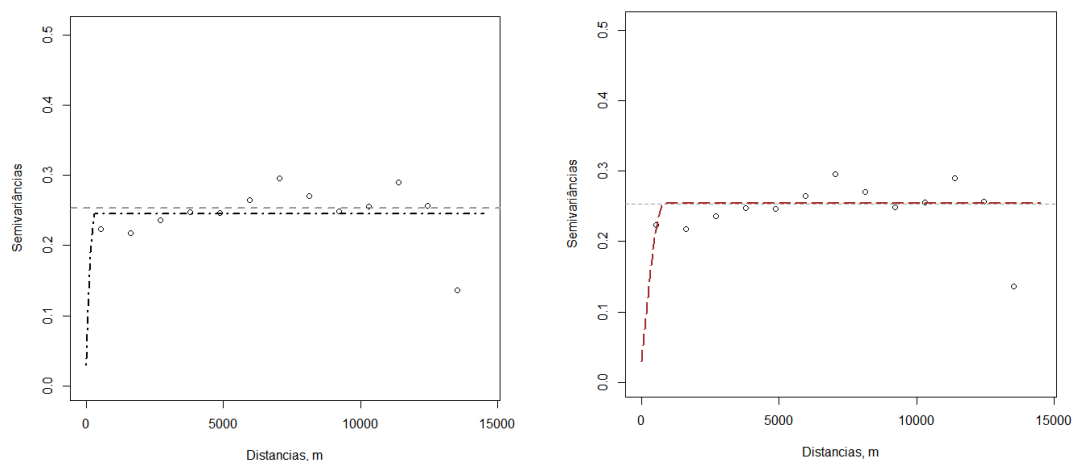


Figura 39: Gráficos dos ajustes dos semivariogramas: à esquerda com Mínimos Quadrados Ordinários (OLS) e à direita com Mínimos Quadrados Ponderados (WLS), para os casos de malária confirmados no ano de 1999 em Porto Velho-RO.

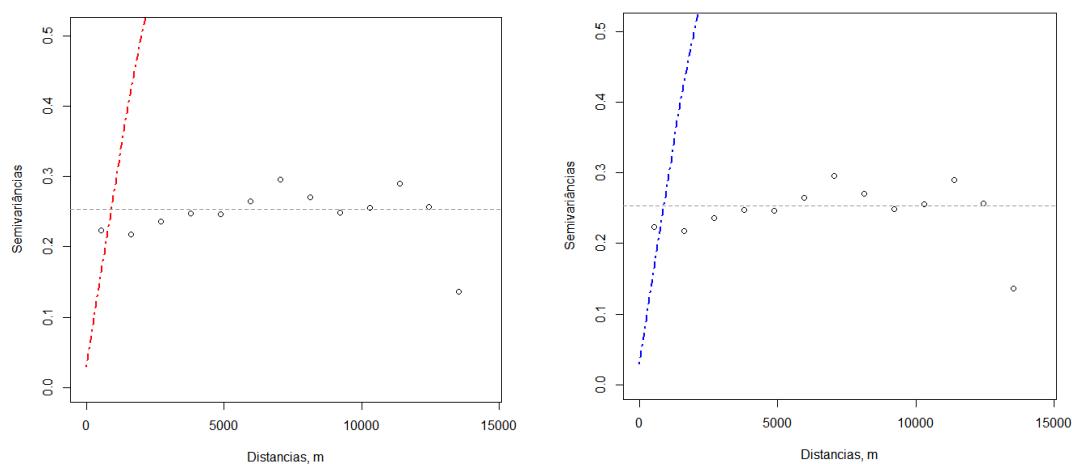


Figura 40: Gráficos dos ajustes dos semivariogramas: à esquerda com a Máxima Verossimilhança (ML) e à direita com a Máxima Verossimilhança Restrita (REML), para os casos de malária confirmados no ano de 1999 em Porto Velho-RO.

Em comparação das Figuras 39 e 40, observamos que os modelos ajustados pelo mínimos quadrados se mostram como o melhor ajuste para o conjunto de dados. Fizemos a autovalidação de ambos os ajustes mínimos quadrados, na Tabela 7:

Tabela 7: Autovalidação dos ajustes Mínimos Quadrados Ordinários (OLS) e Mínimos Quadrados Ponderados (WLS) para os dados de malária.

	OLS	WLS
Média dos erros padronizados	0,00114	0,00131
Desvio dos erros padronizados	1,01638	1,12192
Variância dos erros	0,25602	0,29338
Média dos erros	-0,00110	-0,00082
β_0 da Regressão	1,33109	0,79447
β_1 da Regressão	-1,66583	-0,58990
R^2 ajustado	-0,00314	0,01455
AIC	135,9073	134,3066
BIC	143,4067	141,8060

Observando os resultados apresentados na Tabela 7, o modelo de Mínimos Quadrados Ponderados (WLS) também se mostrou um pouco melhor com o ajuste para o conjunto de dados de malária, tendo valores menores, como o AIC e BIC.

A Tabela 8 mostra os resultados obtidos pelo semivariograma e os valores de variâncias da Krigagem Indicativa para os casos de malária:

Tabela 8: Resultados referentes a modelagem do semivariograma para a Krigagem Indicativa dos casos de malária.

Modelagem do Semivariograma:	
Método de ajuste	Mínimos Quadrados Ponderados (WLS)
Alcance prático	795,9304
Contribuição	0,2240135
Efeito pepita	0,03
Modelo ajustado	esférico

Na Tabela 8, temos um alcance prático alto e um efeito pepita baixo,

sendo um bom ajuste para o conjunto de dados.

4.9 Krigagem Indicativa para os dados de malária

Na Figura 41, temos o mapa de isolinhas referente aos casos de malária, cada valor de isolinha representa a probabilidade de haver casos de malária:

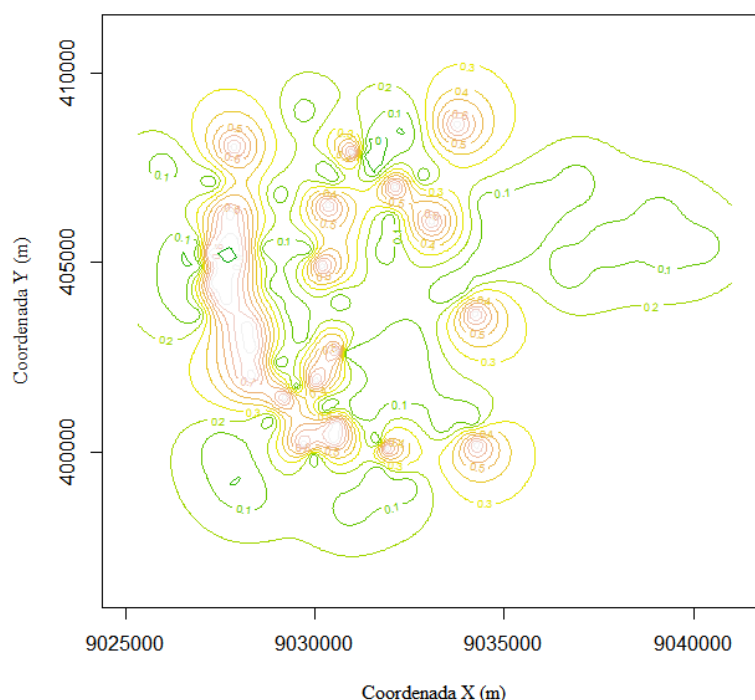


Figura 41: Mapa de isolinhas dos casos de malária confirmados no ano de 1999 em Porto Velho-RO.

No mapa da Krigagem Indicativa, Figura 42, os pontos mais claros mostram os locais com maiores ocorrências. Observa-se que a maior incidência está na região oeste.

No mapa, Figura 43, observamos que houve baixa variância da Krigagem Indicativa em que houve uma variância em torno de 0.08% onde os pontos estavam mais afastados, porém onde os pontos estavam centrados houve uma variação menor, em torno de 0.02%. Os resultados são confirmados pelos dados da Tabela 9.

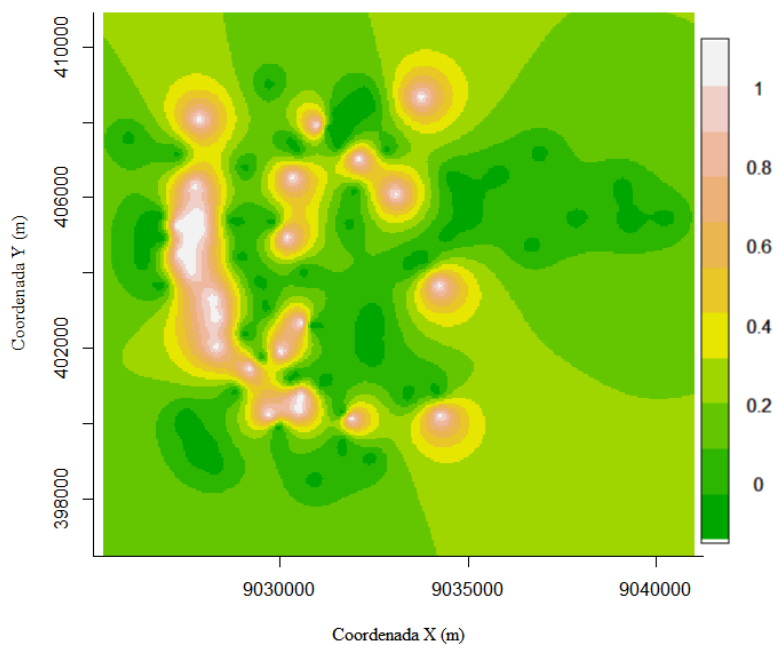


Figura 42: Mapa de Krigagem Indicativa referente aos casos de malária confirmados no ano de 1999 em Porto Velho-RO.

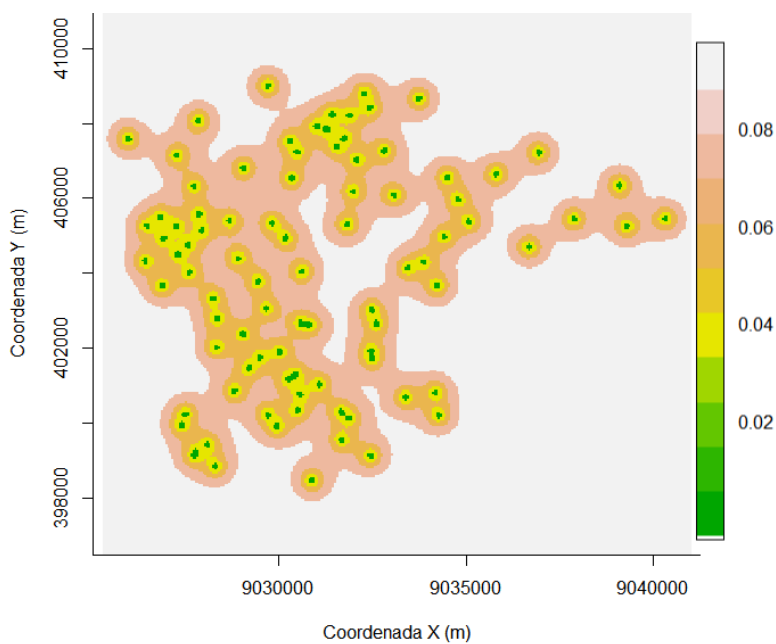


Figura 43: Variância de Krigagem referente aos casos de malária confirmados no ano de 1999 em Porto Velho-RO.

Tabela 9: Variância de Krigagem Indicativa dos casos de malária.

Média	0,1615071
Variância	0,001549266
Coefficiente de variação	0,2437088
Erro padrão	0,004148984

4.10 Ajuste da função K de Ripley para os dados de malária

No gráfico da função K, Figura 44, temos a estimativa simulada pelo valor real da distância para os dados de malária, em que constamos uma concentração dos casos.

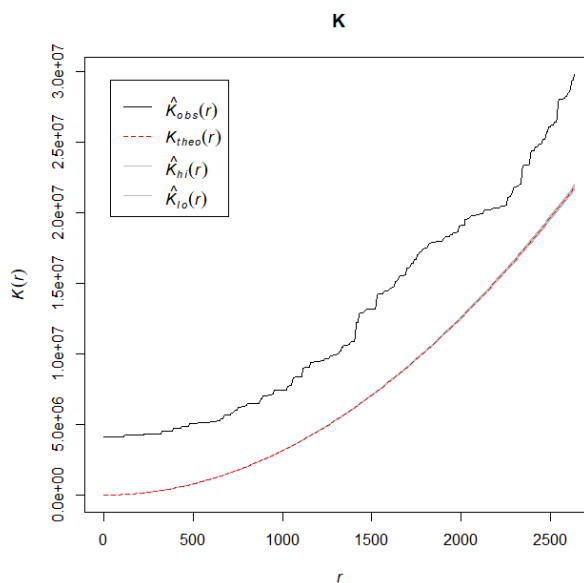


Figura 44: Gráfico da função K (função de Ripley) dos casos de malária confirmados no ano de 1999 em Porto Velho-RO.

No gráfico da função F para os dados de malária, Figura 45, temos a estimação simulada pelo valor real, que indica espaço vazio baixo.

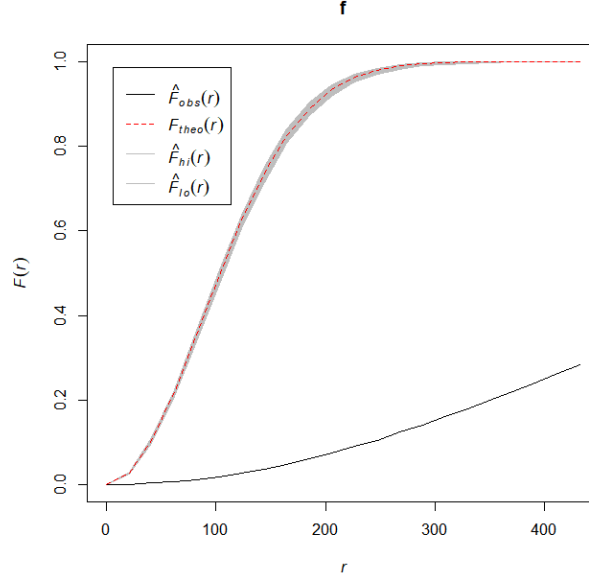


Figura 45: Gráfico da função F (função de espaço vazio ou taxa de risco) dos casos de malária confirmados no ano de 1999 em Porto Velho-RO.

4.11 Kernel para os dados de malária

Assim como para os casos de dengue, para a aplicação do Kernel precisávamos primeiro de um raio de influência, que fosse adequado para a suavização. Pela validação cruzada o raio de influência encontrado foi dado pelo valor de $\sigma = 208,2787$. Como visto no gráfico da Figura 46.

Após ter encontrado o raio de influência fizemos a densidade da estimação Kernel, com o Kernel Gaussiano, gerando a superfície de densidade, Figura 47, onde a região mais clara representa a concentração dos casos de malária.

Observamos a Figura 47, que além de ter mais casos na região oeste, também há casos na região leste do mapa.

O mapa de incerteza, Figura 48, mostra que quanto mais clara a região, maior a incerteza. Neste caso também há algumas dessas incertezas associadas à estes

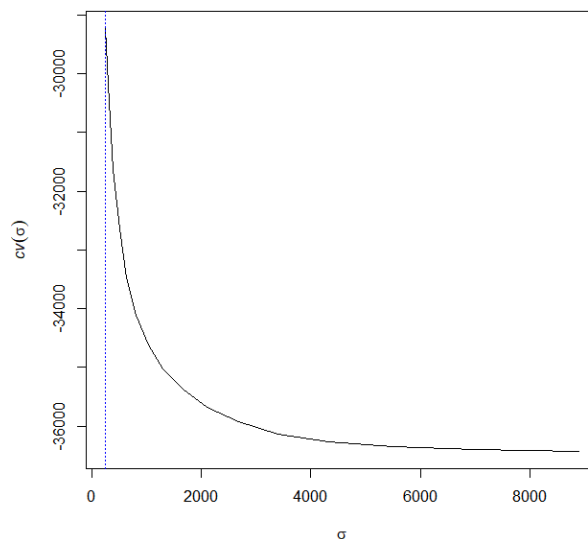


Figura 46: Gráfico do raio de influência, com $\sigma = 208,2787$, para os casos de malária confirmados no ano de 1999 em Porto Velho-RO.

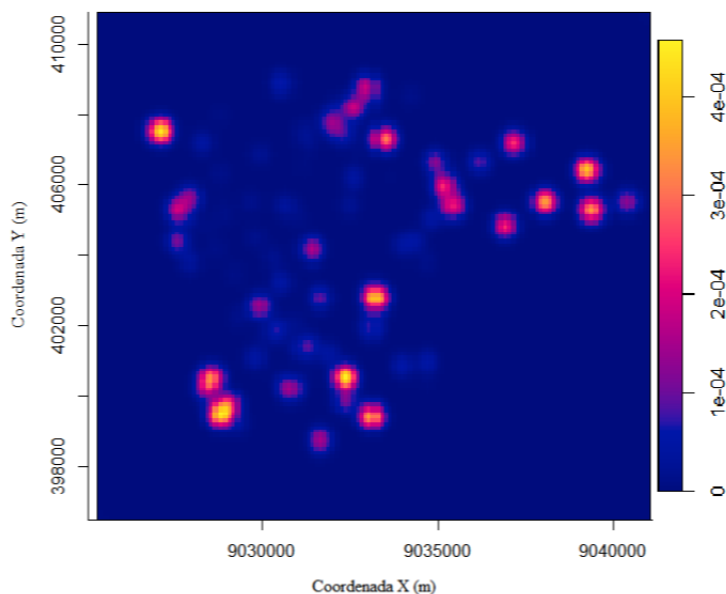


Figura 47: Mapa de estimação da densidade Kernel referente aos casos de malária confirmados no ano de 1999 em Porto Velho-RO.

dados pela resolução utilizada por esse sistema, que não permite distinguir as áreas onde ocorreu e ou não os casos, quando há uma concentração de ocorrências.

Os critérios de informação Akaike (AIC) e Bayesiano (BIC) para es-

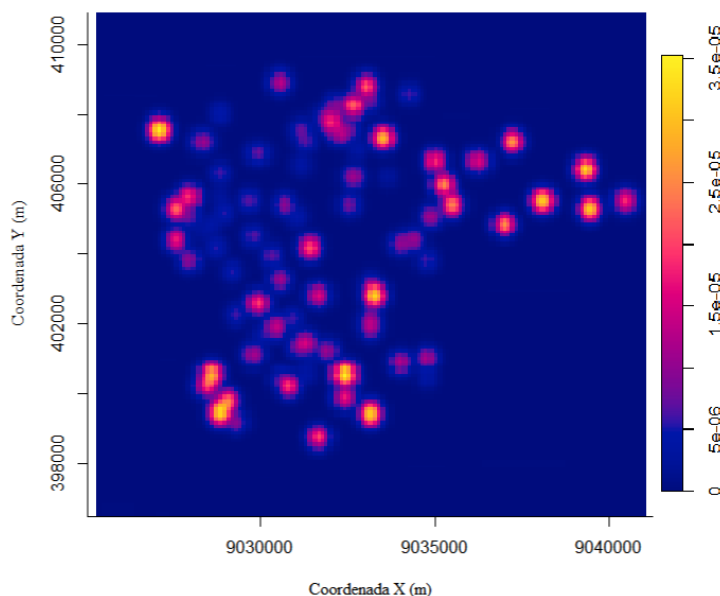


Figura 48: Mapa de incerteza da estimação da densidade Kernel referente aos casos de malária confirmados no ano de 1999 em Porto Velho-RO.

timização da densidade Kernel para os casos de malária, são vistos na Tabela 10.

Tabela 10: Critérios de informação Akaike (AIC) e Bayesiano (BIC) da estimação de densidade Kernel dos casos de malária.

AIC	2,733993
BIC	-1,266007

4.12 Comparação das técnicas Kernel e Krigagem Indicativa para os dados de malária

Comparando os mapas Kernel e Krigagem Indicativa para os dados de malária, observamos na Figura 49 que o mapa Kernel há mais pontos de concentração da endemia do que o mapa da Krigagem Indicativa, que parece ter uma densidade maior somente na região oeste, lembrando que a krigagem representa a probabilidade de o valor de corte ser superado e o kernel representa a intensidade de ocorrência em cada ponto então comparamos os critérios de informação.

Visualmente, o mapa da técnica Kernel se apresentou melhor, pois mostra mais regiões em que há à endemia.

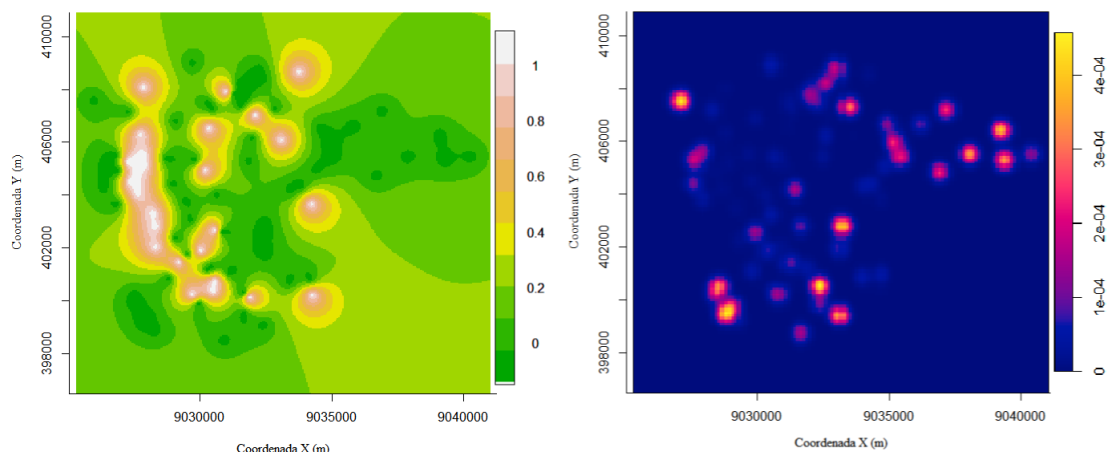


Figura 49: Comparação dos mapas para os dados de malária, á direita o mapa Kernel e á esquerda o mapa da Krigagem Indicativa.

Agora, comparando os critérios de informação, neste caso, o AIC e BIC para os dados de malária, Tabela 11, também mostra que a técnica Kernel se mostrou melhor na precisão das predições em comparação ao Kernel, por apresentar valores mais baixos.

Tabela 11: Critérios de informação Akaike (AIC) e Bayesiano (BIC) da estimação de densidade Kernel e da Krigagem Indicativa dos casos de malária.

	AIC	BIC
Krigagem Indicativa	134,3066	141,8060
Kernel	2,733993	-1,266007

5 CONSIDERAÇÕES FINAIS

O objetivo inicial deste trabalho foi estudar e comparar as técnicas de estatística espacial, o Kernel e a Krigagem Indicativa para conjuntos de dados de dengue e malária. Os objetivos particulares foram aplicar ambas as técnicas para os mesmos conjuntos de dados e calcular os critérios de informação Akaike (AIC) e Bayesiano (BIC) a fim de compará-las e verificar qual delas apresentou melhor precisão nas predições.

Nosso objetivo foi alcançado, pois mostra o potencial de aplicação das técnicas. Além disso, com os dados obtidos pelos mapas é possível verificar as áreas com a concentração das endemias e fazer planejamentos de prevenção e obras públicas que visem na diminuição ou erradicação das mesmas.

Ao longo deste trabalho vimos as vantagens e desvantagens na aplicação das técnicas. A Krigagem Indicativa tem vantagem de modelar atributos mesmo tendo uma alta variabilidade espacial, não precisando eliminar os *outliers* e permite a estimação de incertezas através da função de distribuição acumulada da variável aleatória; e tem a desvantagem de requer um maior conhecimento dos dados em estudo para se definir o ponto de corte para a transformação das variáveis, e também a escolha dos variogramas pode não ser uma tarefa simples. Enquanto o Kernel tem a vantagem não depender dos limites físicos e políticos, e tem desvantagem quando existe uma concentração grande de pontos, prejudicando a análise visual e a escolha do melhor raio de influência pode não ser muito simples.

Em relação as dificuldades que tivemos em fazer as comparações das técnicas, está relacionado à falta de variáveis adicionais nos bancos de dados, no qual nos impôs algumas limitações; com isso o melhor meio de compará-las foi através dos

critérios de informações Akaike (AIC) e Bayesiano (BIC).

Os valores encontrados nos critérios de informação AIC e BIC para ambas as técnicas mostraram condizentes com os valores encontrados na literatura. Nas comparações, a técnica Kernel sobressaiu sobre a técnica Krigagem Indicativa para ambos os conjuntos de dados, o que pode não ocorrer para outros conjuntos de dados. Talvez uma forma melhor para fazer essa comparação seria fazer uma regressão para técnica Kernel, para ter os valores dos critérios de informação mais próximos dos valores da técnica de Krigagem Indicativa. Isso não foi possível neste trabalho devido á falta de informações nos bancos de dados, mas comparando as técnicas por um critério visual, o mapa da técnica Kernel também se mostrou melhor para ambos os conjuntos de dados, pois permite uma melhor visualização das áreas em que houve as endemias.

Ao final concluímos que essas técnicas não são comparáveis, mas sim complementares, pois a técnica Kernel mostrou a intensidade das endemias e a técnica Krigagem Indicativa mostrou a probabilidade desses valores serem maiores que um ponto de corte.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDRADE, A. L.; MONTEIRO, A. M. V.; BARCELLOS, C.; LISBOA, E.; ACOSTA, L. M. W.; ALMEIDA, M. C. M.; BRITO, M. R. V.; CARVALHO, M. S.; SANTOS, M. A.; CRUZ, O.; SANTOS, R. S.; FLORES, R.; SANTOS, S. M.; SILVA, S. A.; SOUZA, W. V. **Introdução à Estatística Espacial para a Saúde Pública**. Brasília: Editora MS, 2007. 120p.

AZEVEDO, T. S.; PIOVEZAN, R.; ZUBEN, C. J. V.; ANDRÉ, I. R. N.; ALMEIDA, D. Perfil epidemiológico da dengue no município de Rio Claro no período de 1996 a 2010. **Hygeia**, v.7, n.12, p.19–30, 2011.

BADDELEY, A.; RUBAK, E.; TURNER, R. **Spatial Point Patterns: Methodology and Applications with R**. London: Chapman and Hall/CRC Press, 2015.

BAGNARA, D.; PRIETTO, P. D. M.; TIMBOLA, R. S. Aplicação da krigagem ordinária na modelagem do pH e da dureza da água subterrânea na área central de Passo Fundo-RS. **Teoria e prática na engenharia civil**, v.22, n.20, p.15–22, 2012.

BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. London: Longman Scientific & Technical Essex, 1995. 413p.

BANNARI, A.; KADHEM, G. MBES–CARIS data validation for bathymetric mapping of shallow water in the Kingdom of Bahrain on the Arabian Gulf. **Remote Sensing**, v.9, n.4, p.1–385, 2017.

BARATA, R. C. B. Malária no Brasil: panorama epidemiológico na última década. **Cadernos de Saúde Pública**, v.11, n.3, p.128–136, 1995.

BARBOSA, G. L. Distribuição espacial dos indicadores entomológicos de *Aedes aegypti* e associação com a ocorrência de casos de dengue em município de médio porte do Estado de São Paulo. Campinas, 2014. 88p. Thesis(Ph.D.) - UNICAMP.

CAMARGO, E. C. G. Desenvolvimento, implementação e teste de procedimentos geoestatísticos (krigagem) no sistema de processamento de informações georreferenciadas (Spring). **São José dos Campos**, 1997.

CAMBARDELLA, C. A.; MOORMAN, T. B.; PARKIN, T. B.; KARLEN, D. L.; NOVAK, J. M.; TURCO, R. F.; KONOPKA, A. E. Field-scale variability of soil properties in central Iowa soils. **Soil science society of America journal**, v.58, n.5, p.1501–1511, 1994.

CLARO, L. B. L.; TOMASSINI, H. C. B.; ROSA, M. L. G. Prevenção e controle do dengue: uma revisão de estudos sobre conhecimentos, crenças e práticas da população. **Cadernos de Saúde Pública**, v.20, p.1447–1457, 2004.

COSTA, J. V.; DONALISIO, M. R.; SILVEIRA, L. V. A. Spatial distribution of dengue incidence and socio-environmental conditions in Campinas, São Paulo State, Brazil, 2007. **Cadernos de Saúde Pública**, v.29, p.1522–1532, 2013.

CRAGLIA, M.; MAHESWARAN, R. **GIS in Public Health Practice**. Boca Raton: CRC Press, 2004. 322p.

CVE. Boletim epidemiológico da dengue no município de Rio Claro em 7 de julho de 2010. <http://www.saude.sp.gov.br/resources/cve-centro-de-vigilancia-epidemiologica/publicacoes/e-becve/e-becven72010.pdf>, 2010. Online: acessado 05 de junho de 2019.

DIAS, J. C. P. Problemas e possibilidades de participação comunitária no controle das grandes endemias no Brasil. **Cadernos de Saúde Pública**, v.14, p.S19–S37, 1998.

DIGGLE, P.; MARRON, J. Equivalence of smoothing parameter selectors in density and intensity estimation. **Journal of the American Statistical Association**, v.83, n.403, p.793–800, 1988.

FERNANDES, T. T. Krigagem Indicativa para Elaboração de Mapas Probabilístico em Agricultura de Precisão. Botucatu, 2014. 67p. Dissertação (Mestrado) - Universidade Estadual Paulista (UNESP).

FERREIRA, M. E. M. C. “Doenças tropicais”: o clima e a saúde coletiva. Alterações climáticas e a ocorrência de malária na área de influência do reservatório de Itaipu, PR. **Terra Livre**, v.1, n.20, p.179–192, 2015.

GIEHL, E. L. H.; BUDKE, J. C.; ATHAYDE, E. A. Distribuição espacial de espécies arbóreas em uma floresta estacional em Santa Maria, sul do Brasil. **Pesquisas: Botânica, São Leopoldo**, v.0, n.58, p.215–226, 2007.

GOVERNO DO ESTADO DE SÃO PAULO - SECRETARIA DE ESTADO DA SAÚDE. SUCEN - Superintendência de Controle de Endemias. <http://www.saude.sp.gov.br/sucen-superintendencia-de-controle-de-endemias/programas/malaria/doenca>, 2019. Online; acessado 26 de agosto de 2019.

GUERRA, P. A. G.; SALLES, J. J. C. **Geoestatística operacional**. Ministério das Minas e Energia, Departamento Nacional da Produção Mineral, 1998. 145p.

HINO, P.; VILLA, T. C. S.; SASSAKI, C. M.; NOGUEIRA, J. A.; SANTOS, C. B. Geoprocessamento aplicado à área da saúde. **Revista Latino-Americana de Enfermagem**, v.14, n.6, 2006.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA-IBGE. Indígenas. https://indigenas.ibge.gov.br/images/indigenas/mapas/pop_indigena_tot_2010.pdf, 2019. Online; acessado 15 de dezembro de 2019.

JOURNEL, A. G. Nonparametric estimation of spatial distributions. **Journal of the International Association for Mathematical Geology**, v.15, n.3, p.445–468, 1983.

JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining geostatistics**. London: Academic press London, 1978. 600p.

LANDIM, P. M. B. **Análise estatística de dados geológicos**. São Paulo: Unesp, 2003. 259p.

LANDIM, P. M. B. Sobre geoestatística e mapas. **Terra e Didática**, v.2, n.1, p.19–33, 2006.

LEROY, O.; GAUVREAU, B.; JUNKER, F.; DE ROCQUIGNY, E.; BERENGIER, M. Uncertainty assessment for outdoor sound propagation. In: INTERNATIONAL CONGRESS ON ACOUSTICS - ICA, 20, Sydney, 2010. **Proceedings of 20th International Congress on Acoustics - ICA**; resumos. Australia: ICA, 2010. 7p.

LOURENÇO, R. W.; LANDIM, P. M. B.; FERREIRA, M. C. Análise da distribuição espacial da produção de monóxido de carbono (CO) em áreas urbanas a partir de superfícies de tendência. **Geografia**, v.26, n.2, p.127–138, 2001.

MARTINS, L. O. Detecção de Massas em Imagens Mamográficas Através do Algoritmo Growing Neural Gas e da Função K De Ripley. São Luís, 2007. 106p. Dissertação (Mestrado) - Universidade Federal do Maranhão.

MATHERON, G. Principles of geostatistics. **Economic geology**, v.58, n.8, p.1246–1266, 1963.

MATSUMOTO, P. S. S.; FLORES, E. F. Estatística espacial na geografia: um estudo dos acidentes de trânsito em Presidente Prudente–SP. **Geografia em Atos**, v.1, n.12, p.95–113, 2012.

MINISTÉRIO DA SAÚDE. Malária: o que é, causas, sintomas, tratamento, diagnóstico e prevenção. <http://www.saude.gov.br/saude-de-a-z/malaria>, 2019a. Online; acessado 26 de agosto de 2019.

MINISTÉRIO DA SAÚDE. Situação Epidemiológica: Dengue, Chikungunya, Zica, SZC, Sarampo e Influenza. <http://portalarquivos2.saude.gov.br/images/pdf/2019/marco/29/3.a-b-Situacao-epidemiologica-da-Dengue-Zika-e-Chikungunya-Sarampo-e-Influenza.pdf>, 2019b. Online; acessado 19 de dezembro de 2019.

MONTEIRO, A. M. V.; CÂMARA, G.; DAVIS, C. Geoestatística: fundamentos e aplicações. In: CAMARGO, E. C. G. (Ed.). **Geoprocessamento: teoria e aplicações**. São José dos Campos: INPE, 2004a. p.5–36.

MONTEIRO, A. M. V.; FELGUEIRAS, C. A.; ASSAD, E.; CAMARGO, E. C. G.; CÂMARA, G.; CARVALHO, M. S.; CRUZ, O. G.; FUCKS, S. D.; CORREA, V.; SOUZA, W. Análise Espacial de Eventos. In: CÂMARA, G.; CARVALHO, M. S. (Ed.). **Análise Espacial de Dados Geográficos**. Brasília: EMBRAPA, 2004b. p.2–15.

MONTEIRO, A. M. V.; FELGUEIRAS, C. A.; ASSAD, E.; CAMARGO, E. C. G.; CÂMARA, G.; CARVALHO, M. S.; CRUZ, O. G.; FUCKS, S. D.; CORREA, V.; SOUZA, W. Análise Espacial e Geoprocessamento. In: CÂMARA, G.; MONTEIRO, A. M. V.; CARVALHO, M. S.; FUCKS, S. D. (Ed.). **Análise Espacial de Dados Geográficos**. Brasília: EMBRAPA, 2004c. p.21–52.

MOURA, A. S.; ROCHA, R. L. Endemias e epidemias: dengue, leishmaniose, febre amarela, influenza, febre maculosa e leptospirose. **Belo Horizonte: Nescon/UFMG**, p.1–82, 2012.

OPAS. Malária. https://www.paho.org/bra/index.php?option=com_content&view=article&id=5287:malaria-2&Itemid=875, 2016. Online; acessado 26 de agosto de 2019.

PEREIRA, P. E. C. Estimativa de recursos minerais e otimização de cava aplicados a um estudo de caso de uma mina de calcário. Catalão, 2017. 171p. Dissertação (Mestrado) - Universidade Federal de Goiás.

PIOVEZAN, R. Levantamento de larvas de Culicidae (Diptera) em diferentes criadouros no município de Santa Bárbara D'Oeste, SP. Rio Claro, 2009. 111p. Dissertação (Mestrado) - Universidade Estadual Paulista (UNESP).

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria, 2018.

RAMÍREZ, J. E. G. Variabilidade Espacial do Parâmetro Geomecânico RQD no Depósito Mineral Animas-Peru. Rio de Janeiro, 2009. 162p. Thesis(Ph.D.) - PUC-Rio.

RIBEIRO JR, P. J.; DIGGLE, P. J. **geoR: Analysis of Geostatistical Data**, 2018. R package version 1.7-5.2.1.

RODRIGUES JR, A. L.; NETTO, A. R.; CASTILHO, E. A. Distribuição espacial do índice de desenvolvimento humano, da infecção pelo HIV e da comorbidade AIDS-tuberculose: Brasil, 1982–2007. **Revista Brasileira de Epidemiologia**, v.17, n.supl. 2, p.204–215, 2014.

ROSSI, M. E.; DEUTSCH, C. V. **Mineral resource estimation**. Florida: Springer Science & Business Media, 2013. 332p.

SANTOS, A. S. Krigagem de teores de ouro da Mina de Caiamar-Goiás. São Paulo, 2015. 78p. Thesis(Ph.D.) - Univerdade de São Paulo.

SIMÃO, F. B. Mapeamento de risco de malária na área urbana de Porto Velho–RO, pela krigagem indicativa. Rio Claro, 2001. 67p. Dissertação (Mestrado) - Universidade Estadual Paulista (UNESP).

SOUZA, L. J. **Dengue—diagnóstico, tratamento e prevenção**. Rio de Janeiro: Editora Rubio, 2008. 200p.

STURARO, J. R. Apostila de geoestatística básica. **Rio Claro, UNESP–IGCE**, v.1, n.1, p.1–34, 2015.

UNIVERSITY COLUMBIA. Population health methods. <https://www.mailman.columbia.edu/research/population-health-methods/kriging>, 2019. Online; acessado 15 de abril de 2019.

VASCONCELOS, V. V. Análise de eventos pontuais. https://pt.slideshare.net/vitor_vasconcelos/anlise-espacial-de-eventos-pontuais, 2016. Online; acessado 25 de agosto de 2019.

WANDERLEY, H. S.; AMORIM, R. F. C.; CARVALHO, F. O. Variabilidade espacial e preenchimento de falhas de dados pluviométricos para o estado de Alagoas. **Revista Brasileira de Meteorologia**, v.27, n.3, p.347–354, 2012.

WIKIPÉDIA. Porto Velho. https://pt.wikipedia.org/wiki/Porto_Velho, 2019. Online; acessado 26 de agosto de 2019.

YAMAMOTO, J. K. Comparação de métodos computacionais para avaliação de reservas: um estudo de caso na jazida de cobre de Chapada, GO. São Paulo, 1991. 172p. Tese de doutoramento - Universidade de São Paulo.

YAMAMOTO, J. K. **Avaliação e classificação de reservas minerais**. São Paulo: EdUSP, 2001. 235p.

YAMAMOTO, J. K.; LANDIM, P. M. B. **Geoestatística: conceitos e aplicações**. São Paulo: Oficina de textos, 2015. 210p.

ZIMBACK, C. R. L. Análise espacial de atributos químicos de solos para fins de mapeamento da fertilidade do solo. Botucatu, 1994. 114p. Tese (Doutorado) - Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista.