

**MARCONI MARTINS CUNHA**

**PROPOSTA METODOLÓGICA PARA A DETERMINAÇÃO DO TAMANHO  
AMOSTRAL E DISTRIBUIÇÃO ESPACIAL DE FEIÇÕES LINEARES UTILIZADAS  
NO CONTROLE DE QUALIDADE POSICIONAL EM CARTOGRAFIA**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Civil, para obtenção do título de *Doctor Scientiae*.

Orientador: Afonso de Paula dos Santos

Coorientadores: Nilcilene das Graças Medeiros  
Marcelo Antônio Nero

**VIÇOSA – MINAS GERAIS**

**2024**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade  
Federal de Viçosa - Campus Viçosa**

T

C972p  
2024

Cunha, Marconi Martins, 1993-  
Proposta metodológica para a determinação do tamanho amostral e distribuição espacial de feições lineares utilizadas no controle de qualidade posicional em Cartografia / Marconi Martins Cunha. – Viçosa, MG, 2024.

1 tese eletrônica, 244 f.: il. (algumas color.).

Inclui apêndices.

Orientador: Afonso de Paula dos Santos.

Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Engenharia Civil, 2024.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2024.129>

Modo de acesso: World Wide Web.

1. Cartografia - Controle de qualidade. 2. Análise espacial. 3. Dados geoespaciais. 4. Distribuição espacial. I. Santos, Afonso de Paula dos, 1984-. II. Universidade Federal de Viçosa. Departamento de Engenharia Civil. Programa de Pós-graduação em Engenharia Civil. III. Título.

CDD 22. ed. 526


**MARCONI MARTINS CUNHA**

**PROPOSTA METODOLÓGICA PARA A DETERMINAÇÃO DO TAMANHO AMOSTRAL E DISTRIBUIÇÃO ESPACIAL DE FEIÇÕES LINEARES UTILIZADAS NO CONTROLE DE QUALIDADE POSICIONAL EM CARTOGRAFIA**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Civil, para obtenção do título de *Doctor Scientiae*.


APROVADA: 26 de março de 2024.

Assentimento:

Documento assinado digitalmente  
 **MARCONI MARTINS CUNHA**  
Data: 05/04/2024 14:36:54-0300  
Verifique em <https://validar.iti.gov.br>

---

Marconi Martins Cunha  
Autor

Documento assinado digitalmente  
 **AFONSO DE PAULA DOS SANTOS**  
Data: 09/04/2024 08:15:52-0300  
Verifique em <https://validar.iti.gov.br>

---

Afonso de Paula dos Santos  
Orientador

*Aos meus pais, irmãos e Branca.*

## **AGRADECIMENTOS**

A Deus, por me dar oportunidades, perseverança, paciência e capacidade de desenvolver este estudo.

A minha família, ainda meus quatros pilares: Maria da Consolação, Sidney, Filipe e Yasmin, por todo o apoio que sempre me deram e me darão. Especialmente aos meus pais por me mostrarem que a maior herança é a educação e não os bens materiais; por me ensinarem a ser uma pessoa de bem.

Aos amigos da salinha: Guilherme, Priscila, Raiane, Larissa e Youlia, pelas discussões, ajuda e pelos momentos de descontração. Dentre estes, um agradecimento especial aos grandes amigos Guilherme Cecatto e Larissa Messias pela parceria, por toda ajuda, pelos conselhos e pelos momentos de diversão em todos os momentos da pós-graduação.

Ao professor Afonso de Paula dos Santos, por todo apoio e incentivo desde a graduação.

Ao professor Marcelo Antônio Nero, por ser um incentivador e ser sempre solícito quando necessário.

À professora Nilcilene das Graças Medeiros, pelos ensinamentos proporcionados desde a graduação.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

“É muito melhor lançar-se em busca de conquistas grandiosas, mesmo expondo-se ao fracasso, do que alinhar-se com os pobres de espírito, que nem gozam muito nem sofrem muito, porque vivem numa penumbra cinzenta, onde não conhecem nem vitória nem derrota”.

**Theodore Roosevelt**

## RESUMO

CUNHA, Marconi Martins, D.Sc., Universidade Federal de Viçosa, março de 2024. **Proposta metodológica para a determinação do tamanho amostral e distribuição espacial de feições lineares utilizadas no controle de qualidade posicional em cartografia.** Orientador: Afonso de Paula dos Santos. Coorientadores: Nilcilene das Graças Medeiros e Marcelo Antônio Nero.

Usualmente, a avaliação da acurácia posicional de produtos cartográficos é realizada por meio das discrepâncias posicionais e técnicas baseadas em pontos. Entretanto, a utilização de feições lineares apresenta algumas vantagens sobre o método pontual. Dentre estas, pode-se mencionar que as feições lineares possuem mais informações geométricas e posicionais que os pontos e que, em uma base cartográfica, aproximadamente 80% das feições são linhas. Apesar destas vantagens, parâmetros importantes para a avaliação da acurácia utilizando linhas não foram estabelecidos ou determinados, como o tamanho da amostra ou o seu padrão de distribuição espacial. A distribuição espacial é um fator relevante, pois pode interferir nos resultados e determinar a validade de um processo de avaliação. Diante disso, em seu primeiro capítulo, este estudo propôs um método baseado na modificação do Método do Vizinho Mais Próximo para pontos para avaliar o padrão de distribuição espacial de feições lineares. O método proposto, denominado de Nearest Neighbor Method for Linear Features (NNMLF), foi aplicado a dados simulados e reais. Os resultados mostraram que o NNMLF foi eficaz em estimar o padrão de distribuição espacial esperado em todos os experimentos com dados simulados. Já a aplicação nos dados reais mostrou que o NNMLF é de simples utilização. Com relação ao tamanho amostral, para o controle posicional utilizando linhas, não se tem conhecimento de estudo que indique o tamanho da amostra que deve ser utilizado. Diante disso, em seu segundo capítulo, este estudo propôs um método para determinar este tamanho amostral baseado na Simulação Monte Carlo. Para isso, foi utilizada uma base de dados composta por 46460 km de estradas de 10 províncias no leste da Espanha, e analisado os tamanhos de área de 1; 2,5; 5; 7,5; 10; 15; 20; 25; 50; 75; 100; 125; 150; 175; 200; 225 e 250 km<sup>2</sup>. Por fim, a comparação entre as discrepâncias fornecidas pelas amostras e pela população foi realizada com o teste Kolmogorov-Smirnov. Desta forma, foi possível realizar uma proposta do tamanho amostral de feições lineares utilizadas na avaliação da acurácia posicional, levando em consideração o

padrão de distribuição espacial da amostra e o tamanho da área avaliada. Esta proposta foi realizada em forma de equação, que fornece o tamanho amostral com base no tamanho da área avaliada. O modelo ajustado indica uma forte associação ( $R^2 = 0,9252$ ) entre o tamanho da área e o tamanho amostral mínimo necessário. Com o desenvolvimento destes métodos foi possível propor uma metodologia geral para a avaliação da acurácia posicional planimétrica realizada com feições lineares, segundo a norma brasileira. Essa metodologia, apresentada no terceiro capítulo deste estudo, compreende: determinação do tamanho amostral, análise do padrão de distribuição espacial, detecção de *outlier*, e análises de tendência e de precisão. Essa metodologia foi utilizada em três bases cartográficas: de pequena, média e grande escala e os resultados comparados com a avaliação resultante do método tradicional, realizado com pontos. Para todas as escalas, a metodologia proposta apresentou resultados compatíveis, ou mais rigorosos, que o método de feições pontuais, demonstrando a viabilidade de sua utilização.

Palavras-chave: Padrão de Distribuição Espacial. Tamanho Amostral. Cartografia.

## ABSTRACT

CUNHA, Marconi Martins, D.Sc., Universidade Federal de Viçosa, March, 2024. **Methodological proposal for determining the sample size and spatial distribution of linear features used in positional quality control in cartography.** Adviser: Afonso de Paula dos Santos. Co-advisers: Nilcilene das Graças Medeiros and Marcelo Antônio Nero.

Usually, the positional accuracy of cartographic products is assessed using positional discrepancies and point-based techniques. However, the use of linear features has some advantages over the point method. These include the fact that linear features have more geometric and positional information than points and that, in a cartographic base, approximately 80% of the features are lines. Despite these advantages, important parameters for assessing accuracy using lines have not been established or determined, such as the size of the sample or its spatial distribution pattern. Spatial distribution is a relevant factor, as it can affect the results and determine the validity of an evaluation process. Therefore, in its first chapter, this study proposed a method based on a modification of the Nearest Neighbor Method for points to evaluate the spatial distribution pattern of linear features. The proposed method, called the Nearest Neighbor Method for Linear Features (NNMLF), was applied to simulated and real data. The results showed that the NNMLF was effective in estimating the expected spatial distribution pattern in all the experiments with simulated data. The application to real data showed that the NNMLF is simple to use. With regard to sample size, for positional control using lines, there is no known study that indicates the sample size that should be used. Therefore, in its second chapter, this study proposed a method for determining this sample size based on Monte Carlo Simulation. To do this, a database consisting of 46460 km of roads from 10 provinces in eastern Spain was used, and the area sizes of 1; 2.5; 5; 7.5; 10; 15; 20; 25; 50; 75; 100; 125; 150; 175; 200; 225 and 250 km<sup>2</sup> were analyzed. Finally, the discrepancies between the samples and the population were compared using the Kolmogorov-Smirnov test. In this way, it was possible to make a proposal for the sample size of linear features used to assess positional accuracy, taking into account the spatial distribution pattern of the sample and the size of the area assessed. This proposal was made in the form of an equation, which provides the sample size based on the size of the area assessed. The adjusted model indicates a strong association ( $R^2 = 0.9252$ ) between the size of the area and

the minimum sample size required. By developing these methods, it was possible to propose a general methodology for assessing planimetric positional accuracy using linear features, in accordance with the Brazilian standard. This methodology, presented in the third chapter of this study, includes: determining the sample size, analyzing the spatial distribution pattern, outlier detection, and trend and precision analyses. This methodology was used on three map bases: small, medium and large scale, and the results were compared with the assessment resulting from the traditional method, carried out using points. For all scales, the proposed methodology showed compatible or more rigorous results than the point feature method, demonstrating the viability of its use.

Keywords: Spatial Distribution Pattern. Sample size. Cartography.

## ÍNDICE DE FIGURAS

### CAPÍTULO 1

|   |    |
|---|----|
| Figura 1 – Padrão de distribuição espacial teórico para pontos (a) e linhas (b).....  | 36 |
| Figura 2 - Menores distâncias Euclidianas dos vértices da linha <i>i</i> para qualquer segmento da linha 1.....   | 38 |
| Figura 3 - Relação do índice R com o padrão de distribuição espacial de feições lineares. .   | 42 |
| Figura 4 – Plugin Nearest Neighbor Method for Linear Features (NNMLF).....  | 43 |
| Figura 5 - Fluxograma do <i>plugin</i> NNMLF.....   | 44 |
| Figura 6 - Primeiro conjunto de linhas simuladas. Padrões agrupado (a), aleatório (b) e disperso (c).....   | 46 |
| Figura 7 - Segundo conjunto de linhas simuladas. (a) Primeiro subconjunto e (b) segundo subconjunto.....  | 48 |
| Figura 8 - Área de Estudo do estudo de caso do NNMLF.....   | 51 |
| Figura 9 - Exemplo de como o aumento do número de feições pode afetar a Distância de Hausdorff. Feição dividida em duas (a), três (b) e quatro (c) partes iguais..... | 56 |
| Figura 10 – Exemplo da aplicação de distância entre as linhas divididas. (a) Linha dividida sem espaçamento e (b) linha dividida e espaçada.....                      | 58 |
| Figura 11 - Exemplo da relação entre o número de feições e o padrão de distribuição. (a) Doze linhas e (b) quatro linhas.....   | 61 |

### CAPÍTULO 2

|   |     |
|---|-----|
| Figura 1 - Método do <i>Buffer</i> Simples.....   | 74  |
| Figura 2 – Porcentagem de inclusão da linha de teste em relação à largura do <i>buffer</i> . (a) Aumento da largura do <i>buffer</i> e (b) distribuição de probabilidade.....   | 75  |
| Figura 3 – Método do <i>Buffer</i> Duplo.....   | 76  |
| Figura 4 - Áreas formadas pela interseção dos <i>bufers</i> das linhas de referência (Q) e de teste (X) no método do <i>Buffer</i> Duplo.....   | 77  |
| Figura 5 - Relação entre as áreas normalizadas e a largura dos <i>buffers</i> .....   | 78  |
| Figura 6 – Método da Distância de Hausdorff.....  | 80  |
| Figura 7 – Método da Influência do Vértice. (a) Comprimento dos segmentos adjacentes de um vértice; e (b) Componentes $\Delta X$ e $\Delta Y$ da distância euclidiana.....  | 81  |
| Figura 8 - Área de Estudo.....  | 84  |
| Figura 9 - Metodologia para obtenção do identificador único.....  | 86  |
| Figura 10- Simulação Monte Carlo.....   | 88  |
| Figura 11 - Influência da largura da caixa no histograma.....   | 93  |
| Figura 12 – Populações proporcionais aos tamanhos máximos das amostras. (a) 1 km <sup>2</sup> ; (b) 2,5 km <sup>2</sup> ; (c) 5 km <sup>2</sup> ; (d) 7,5 km <sup>2</sup> ; (e) 10 km <sup>2</sup> ; (f) 15 km <sup>2</sup> ; (g) 20 km <sup>2</sup> ; (h) 25 km <sup>2</sup> ; (i) 50 km <sup>2</sup> ; (j) 75 km <sup>2</sup> ; (k) 100 km <sup>2</sup> ; (l) 125 km <sup>2</sup> ; (m) 150 km <sup>2</sup> ; (n) 175 km <sup>2</sup> ; (o) 200 km <sup>2</sup> ; (p) 225 km <sup>2</sup> e (q) 250 km <sup>2</sup> ..... | 95  |
| Figura 13 - Discrepâncias posicionais da base de dados. (a) <i>Buffer</i> Simples; (b) <i>Buffer</i> Duplo; (c) Distância de Hausdorff e (d) Influência do Vértice.....   | 99  |
| Figura 14 - Influência do algoritmo Douglas-Peucker no método Distância de Hausdorff. .   | 100 |
| Figura 15 - Influência do algoritmo Douglas-Peucker no método Influência do Vértice.....  | 101 |
| Figura 16 – Média do RMS para cada tamanho amostral na área de 5 km <sup>2</sup> . (a) <i>Buffer</i> Simples; (b) <i>Buffer</i> Duplo; (c) Distância de Hausdorff e (d) Influência do Vértice.....  | 104 |
| Figura 17 - Relação entre tamanho amostral e área.....  | 108 |
| Figura 18 - Exemplo de amostra.....   | 109 |

### **CAPÍTULO 3**

|   |     |
|---|-----|
| Figura 1 – Parte da base de dados da Catalunha. ....  | 122 |
| Figura 2 - Base cartográfica contínua do Rio de Janeiro.....  | 123 |
| Figura 3 - Estradas da Austrália.....   | 124 |
| Figura 4 - Metodologia. ....  | 125 |
| Figura 5 – Análise da tendência (translação) utilizando a Média Direcional e a Variância Circular. (a) Vetores na mesma direção, (b) vetores em direções opostas e (c) vetores com tendência radial. ....                     | 128 |
| Figura 6 - Efeito de escala a partir da Análise Vetorial. (a) Sem efeito de escala e (b) com efeito de escala.....  | 129 |
| Figura 7 - Ângulo de rotação.....   | 130 |
| Figura 8 - Amostras pontuais e lineares utilizadas. (a) Catalunha, (b) Rio de Janeiro e (c) Austrália.....  | 134 |
| Figura 9 - Discrepâncias obtidas pelo método baseado em pontos (Catalunha). ....  | 136 |
| Figura 10 - Discrepâncias obtidas pelo método baseado em linhas (Catalunha). ....   | 136 |
| Figura 11 - Discrepâncias obtidas pelo método baseado em pontos (Rio de Janeiro). ....  | 137 |
| Figura 12 - Discrepâncias obtidas pelo método baseado em linhas (Rio de Janeiro).....   | 137 |
| Figura 13 - Discrepâncias obtidas pelo método baseado em pontos (Austrália). ....   | 138 |
| Figura 14 - Discrepâncias obtidas pelo método baseado em linhas (Austrália).....  | 138 |
| Figura 15 – Vetores de discrepância posicional. (a) Grande escala – pontos, (b) Média escala – pontos, (c) Pequena escala – pontos, (d) Grande escala – linhas, (e) Média escala – linhas, e (f) pequena escala - linhas..... | 139 |
| Figura 16 - Comparação das discrepâncias entre os métodos pontual e linear. (a) Discrepância pontual pequena e (b) Discrepância pontual grande.....   | 141 |

## ÍNDICE DE TABELAS

### CAPÍTULO 1

|   |    |
|---|----|
| Tabela 1 - Constantes para o cálculo do vizinho mais próximo, em função da ordem.....                                   | 40 |
| Tabela 2 - Resultado do NNMLF de Primeira Ordem para dados simulados. ....  | 47 |
| Tabela 3 - Resultado do NNMLF de Segunda e Terceira Ordens para dados simulados.....                                    | 49 |
| Tabela 4 - Distâncias médias entre as linhas, em metros. ....   | 52 |
| Tabela 5 - Vizinho Mais Próximo para Feições Lineares até Terceira Ordem. ....  | 53 |
| Tabela 6 - Valores de $R_{OBS}$ , $R_{ESP}$ e Índice R .....  | 54 |
| Tabela 7 - Valores de Z.....  | 54 |
| Tabela 8 - Influência do número de feições no padrão de distribuição espacial. ....                                     | 55 |
| Tabela 9 - Influência do número de feições no padrão de distribuição espacial, com espaçamento entre os segmentos. .... | 59 |

### CAPÍTULO 2

|   |     |
|---|-----|
| Tabela 1 - Características de cada população. ....  | 97  |
| Tabela 2 - Estatísticas das discrepâncias da base de dados. ....                                  | 98  |
| Tabela 3 - Características das amostras.....  | 103 |
| Tabela 4 – Estatística D para o método <i>Buffer Simples</i> . ....                               | 105 |
| Tabela 5 - Estatística D para o método <i>Buffer Duplo</i> . ....                                 | 106 |
| Tabela 6 - Estatística D para o método Distância de Hausdorff. ....                               | 106 |
| Tabela 7 - Estatística D para o método Influência do Vértice. ....                                | 107 |
| Tabela 8 - Tamanho amostral mínimo representativo da população, por tamanho de área e método..... | 108 |

### CAPÍTULO 3

|  |     |
|--|-----|
| Tabela 1 - Valores de PEC e EP para a planimetria segundo o Decreto nº89.817/84.....                 | 120 |
| Tabela 2 - Valores de PEC e EP para a planimetria.....   | 121 |
| Tabela 3 - Tamanho amostral para a avaliação da acurácia posicional utilizando pontos e linhas. .... | 133 |
| Tabela 4 - Resultado do Método do Vizinho mais Próximo para pontos.....                              | 134 |
| Tabela 5 - Resultado do Nearest Neighbor Method for Linear Features (NNMLF). ....                    | 135 |
| Tabela 6 – Estatísticas descritivas das discrepâncias da área da Catalunha. ....                     | 136 |
| Tabela 7 – Estatísticas descritivas das discrepâncias da área do Rio de Janeiro. ....                | 137 |
| Tabela 8 – Estatísticas descritivas das discrepâncias da área da Austrália. ....                     | 138 |
| Tabela 9 - Análise de tendência para as amostras pontuais. ....                                      | 142 |
| Tabela 10 - Análise de tendência para as amostras lineares. ....                                     | 142 |
| Tabela 11 - Análise da precisão para os dados da Catalunha. Escala de avaliação: 1/5.000. ....       | 144 |
| Tabela 12 - Análise da precisão para os dados do Rio de Janeiro. Escala de avaliação: 1/250.000..... | 144 |
| Tabela 13 - Análise da precisão para os dados da Austrália. Escala de avaliação: 1/5.000.000. ....   | 144 |
| Tabela 14 - Classificação dos produtos segundo o Decreto nº 89.817 e a ET-CQDG. ....                 | 145 |

## SUMÁRIO

|   |           |
|---|-----------|
| <b>AGRADECIMENTOS .....</b>   | <b>4</b>  |
| <b>RESUMO.....</b>  | <b>6</b>  |
| <b>ABSTRACT.....</b>  | <b>8</b>  |
| <b>INTRODUÇÃO GERAL.....</b>  | <b>15</b> |
| 1 Considerações Iniciais .....  | 15        |
| 2 Caracterização do Problema .....  | 17        |
| 3 Justificativa.....  | 24        |
| 4 Hipóteses .....   | 25        |
| 5 Objetivos .....   | 25        |
| 5.1 Objetivo Geral.....   | 25        |
| 5.2 Objetivos Específicos.....  | 25        |
| 6 Estruturação do Trabalho .....  | 26        |
| Referências Bibliográficas .....  | 27        |
| <b>CAPÍTULO 1. PROPOSTA DE UM MÉTODO PARA A AVALIAÇÃO DO PADRÃO DE DISTRIBUIÇÃO ESPACIAL DE FEIÇÕES LINEARES.....</b>   | <b>33</b> |
| <b>1 Introdução .....</b>   | <b>33</b> |
| <b>2 Método Proposto .....</b>  | <b>35</b> |
| <b>3 Implementação de uma Aplicação Computacional para a Utilização do NNMLF.....</b>   | <b>42</b> |
| <b>4 Experimentos e Resultados .....</b>  | <b>45</b> |
| 4.1 Experimento 1: Dados Simulados e as Ordens do NNMLF .....   | 45        |
| 4.2 Experimento 2: Dados Reais em um Estudo de Caso .....   | 50        |
| 4.3 Experimento 3: Dados Reais na Análise da Relação entre o Número de Feições Lineares e o Padrão de Distribuição Espacial.....                                | 54        |
| <b>5 Conclusão e Considerações Finais.....</b>  | <b>62</b> |
| Referências Bibliográficas .....  | 63        |
| <b>CAPÍTULO 2. DETERMINAÇÃO DO TAMANHO AMOSTRAL DE FEIÇÕES LINEARES UTILIZADAS NO CONTROLE DE QUALIDADE CARTOGRÁFICA POR MEIO DA SIMULAÇÃO MONTE CARLO.....</b> | <b>70</b> |
| <b>1 Introdução .....</b>   | <b>70</b> |
| <b>2 Métodos para Avaliação da Acurácia Posicional baseados em Feições Lineares ....</b>  | <b>73</b> |
| 2.1 <i>Buffer</i> Simples .....   | 73        |
| 2.2 <i>Buffer</i> Duplo .....   | 76        |
| 2.3 Distância de Hausdorff .....  | 79        |
| 2.4 Influência do Vértice .....   | 81        |
| <b>3 Simulação Monte Carlo .....</b>  | <b>82</b> |
| <b>4 Materiais e Métodos.....</b>   | <b>84</b> |
| 4.1 Área de Estudo.....   | 84        |
| 4.2 Base de Dados.....  | 84        |

|   |            |
|---|------------|
| 4.3 Metodologia.....  | 85         |
| <b>5 Resultados e Discussões.....</b>   | <b>94</b>  |
| <b>6 Conclusão .....</b>  | <b>109</b> |
| <b>Referências Bibliográficas .....</b>   | <b>110</b> |
| <b>CAPÍTULO 3. PROPOSTA DE UMA METODOLOGIA PARA AVALIAÇÃO DA ACURÁCIA POSICIONAL UTILIZANDO FEIÇÕES LINEARES: O CASO DA NORMA BRASILEIRA E EXEMPLOS DE APLICAÇÃO.....</b> | <b>117</b> |
| <b>1 Introdução .....</b>   | <b>117</b> |
| <b>2 Normas Brasileiras para a Avaliação da Acurácia Posicional.....</b>  | <b>119</b> |
| <b>3 Materiais e Métodos.....</b>   | <b>121</b> |
| 3.1 Base de Dados.....  | 121        |
| 3.2 Metodologia Proposta.....   | 124        |
| <b>4 Resultados e Discussões.....</b>   | <b>131</b> |
| <b>5 Conclusão .....</b>  | <b>145</b> |
| <b>Referências Bibliográficas .....</b>   | <b>147</b> |
| <b>Conclusões Gerais .....</b>  | <b>151</b> |
| <b>Apêndices .....</b>  | <b>152</b> |
| Capítulo 1.....   | 152        |
| Apêndice A. Nearest Neighbor Method for Linear Features (NNMLF) .....   | 152        |
| Capítulo 2.....   | 159        |
| Apêndice A. <i>Buffer</i> Simples.....  | 159        |
| Apêndice B. <i>Buffer</i> Duplo .....   | 165        |
| Apêndice C. Distância de Hausdorff.....   | 174        |
| Apêndice D. Influência do Vértice.....  | 177        |
| Apêndice E. Seleção da Amostra .....  | 181        |
| Apêndice F. Simulação Monte Carlo .....   | 184        |
| Apêndice G. Estatísticas das Amostras.....  | 191        |
| Apêndice H. Gráficos das Discrepâncias das Populações.....  | 192        |
| Apêndice I. Estatísticas das Discrepâncias das Populações .....   | 209        |
| Apêndice J. Relação entre o RMS e o Tamanho Amostral .....  | 212        |
| Capítulo 3.....   | 227        |
| Apêndice A. Método TRS .....  | 227        |

## INTRODUÇÃO GERAL

### 1 Considerações Iniciais

No mundo atual há uma crescente produção de informações em todas as áreas do conhecimento. A evolução tecnológica foi uma das principais responsáveis, senão a principal, por alavancar este crescimento em uma sociedade que cada vez mais demanda informações.

Obviamente, esta revolução também atingiu a Cartografia. Segundo Santos (2015), cada vez mais a informação necessita da componente espacial. De acordo com Lunardi, Penha e Cerqueira (2012), com o crescimento de produtores e usuários das informações espaciais, uma das necessidades que surge é atestar a qualidade dos produtos gerados. Um exemplo deste panorama, que evidencia essa necessidade, é visto em soluções muito populares desde a democratização de informações cartográficas na *internet* (como, por exemplo, Google Earth, Open Street Map, etc.), que são amplamente utilizadas, porém a qualidade dos seus dados é desconhecida e/ou não quantificada e não informada aos seus usuários (ARIZA-LÓPEZ et al., 2020).

Nesse contexto, avaliar um produto cartográfico para se obter conhecimento sobre a sua qualidade e, assim, poder utilizá-lo adequadamente é objetivo principal do Controle de Qualidade Cartográfica (CQC). Essa avaliação é realizada sobre os elementos de qualidade. Segundo Mozas-Calvache e Ariza-López (2011), ISO (2013) e DSG (2016), a completude, a consistência lógica, a acurácia posicional, a acurácia temporal, a acurácia temática e a usabilidade são os principais elementos para caracterizar a qualidade de um produto cartográfico.

Dentre esses elementos, a acurácia posicional é um dos mais estudados e um dos mais importantes (SANTOS, 2015). Devido a isso, a acurácia posicional é o elemento do CQC mais utilizado para verificar a qualidade dos dados espaciais (JAKOBSSON e VAUGLIN, 2002).

Classicamente, essa avaliação da acurácia posicional é realizada pela comparação estatística das discrepâncias posicionais entre a localização de uma entidade espacial no produto geoespacial avaliado e sua localização de referência, de

melhor acurácia (ANSI, 1998; ARIZA-LÓPEZ, RUIZ-LENDÍNEZ e UREÑA-CÁMARA, 2018; CUNHA et. al, 2019). Sendo que, geralmente, essa verificação é realizada por meio de técnicas baseadas em pontos através da comparação entre pontos homólogos bem identificados nos produtos de teste e de referência (FGDC, 1998; VAN NIEL e MCVICAR, 2001; GALO, DAL POZ e FERREIRA, 2001; NERO, 2005, MOZAS-CALVACHE, 2007; MOZAS-CALVACHE e ARIZA-LÓPEZ, 2011; ARIZA-LÓPEZ e MOZAS-CALVACHE, 2012; ARIZA-LÓPEZ et al., 2012; SANTOS et al., 2015; NERO et al., 2017; ARIZA-LÓPEZ, RODRÍGUEZ-AVI e ALBA-FERNÁNDEZ, 2018; CUNHA et al., 2019, MOZAS-CALVACHE, 2021).

Contudo, o avanço da tecnologia, especialmente o surgimento do GNSS (*Global Navigation Satellite System*), ampliou a possibilidade de se utilizar linhas para o controle posicional (MOZAS-CALVACHE, 2007). Sendo que, nas últimas décadas diversos métodos baseados em feições lineares foram propostos (SANTOS et al., 2015). São exemplos os métodos Banda Épsilon (BLAKEMORE ,1984), Distância de Hausdorff (ABBAS, 1994), *Buffer* Simples (GOODCHILD e HUNTER, 1997), *Buffer* Duplo (TVEITE e LANGAAS, 1999), Banda de Erro Simulada Estatisticamente (SSE-Band) (TONG et al., 2013), Influência do Vértice (MOZAS-CALVACHE e ARIZA-LÓPEZ, 2011), Influência do Vértice por Vértices (MOZAS-CALVACHE e ARIZA-LÓPEZ, 2018), entre outros.

Comparado ao método tradicional que utiliza pontos, a utilização de feições lineares para a avaliação da acurácia posicional apresenta algumas vantagens:

- (i) Aproximadamente 80% das entidades de uma base cartográfica são constituídas de feições lineares (CUENIN, 1972);
- (ii) As feições lineares geralmente apresentam boa distribuição espacial nas bases cartográficas, como em redes rodoviárias (MOZAS-CALVACHE e ARIZA-LÓPEZ, 2011);
- (iii) Os elementos lineares possuem variadas informações geométricas, contando vértices, ângulos, comprimento dos segmentos de reta, orientação, sinuosidade, etc. (MOZAS-CALVACHE e ARIZA-LÓPEZ, 2011);
- (iv) As feições lineares são fáceis de serem identificadas e extraídas na imagem, tanto de maneira visual quanto automática (KUBIK, 1991; HABIB et al., 2004); e

- (v) Os métodos baseados em feições lineares são uma boa alternativa em situações em que o número de pontos bem identificáveis é insuficiente ou a sua distribuição espacial não é adequada (GALO, DAL POZ e FERREIRA, 2001). Isso porque há a possibilidade de se utilizar as feições lineares como complemento aos métodos baseados em pontos, não somente como uma substituição (ARIZA-LÓPEZ e MOZAS-CALVACHE, 2012).

Contudo, apesar dessas vantagens, no Brasil ainda não é comum a utilização de feições lineares para a realização do controle de qualidade cartográfica (SANTOS, 2015). Além disso, quando se trata das feições pontuais, já existem normas para a determinação do tamanho amostral e o padrão de distribuição dos elementos de controle. Entretanto, para a avaliação da acurácia posicional utilizando linhas, estes parâmetros, que são de grande importância para o controle posicional, ainda não foram determinados (ARIZA-LÓPEZ, RUIZ-LENDÍNEZ e UREÑA-CÁMARA, 2018).

## **2 Caracterização do Problema**

Na avaliação da acurácia posicional utilizando feições lineares, temas importantes como o tamanho amostral ou o seu padrão de distribuição espacial parecem ter sido pouco ou nada explorados em alguns aspectos. Essa questão pendente é comprovada por meio de uma análise na literatura.

Mozas-Calvache (2007) afirma que, no caso de elementos lineares, não há recomendações sobre a amostragem a ser realizada para a seleção das linhas. Além disso, este autor argumenta que, apesar de existência dos métodos de feições lineares e da sua utilização, em nenhum deles há recomendações sobre amostragem e, portanto, o desenvolvimento de uma metodologia para a determinação da amostra para elementos lineares é um estudo pendente.

Por sua vez, Mozas-Calvache e Ariza-López (2010) também afirmam que não há tamanho amostral recomendado para o caso em que o controle posicional é realizado mediante o uso de feições lineares.

Ariza-López e Mozas-Calvache (2012) destacam que o tamanho da amostra e a distribuição da amostra são aspectos importantes para os métodos de avaliação da

acurácia posicional. Entretanto, também pontuam que estes ainda não foram estabelecidos ou determinados, em termos de elementos lineares.

Já Santos (2015) afirma que não há consenso sobre o tamanho amostral para feições lineares utilizadas no controle de qualidade cartográfica. Afirmação semelhante é realizada por La Vega, Ariza-López e Mozas-Calvache (2016) e por Santos et al. (2016).

Segundo Moreira et al. (2018), considerando o fato da utilização de feições lineares no controle de qualidade cartográfica, atualmente não existe norma ou padrão que definem tamanho amostral para a avaliação da acurácia posicional, bem como padrões de distribuição espacial.

Fonseca (2018) ainda recomenda o desenvolvimento de estudos que possibilitem analisar a distribuição espacial e tamanho amostral de feições lineares.

Yates, Brook e Buettel (2022) propuseram diversas funções para análise de feições lineares em ecologia, mais especificamente para troncos de árvores caídos. Dentre estas funções é proposta uma métrica para medir a densidade de linhas próximas. Entretanto, é considerado que sempre as linhas vizinhas são aleatórias. Os próprios autores afirmam que seria interessante generalizar os padrões, para que o método se assemelhasse às estatísticas tradicionais, como o método do vizinho mais próximo, que analisa diferentes padrões de distribuição espacial (disperso, aleatório e agrupado) e em diferentes ordens.

Durante essa pesquisa bibliográfica se teve conhecimento somente de um trabalho que aborda a questão do tamanho amostral em feições lineares. Ariza-López et al. (2011), utilizando a Simulação Monte Carlo, discutem a influência do tamanho da amostra na variabilidade dos resultados de alguns métodos de feições lineares.

Entretanto, os autores assumem que a distribuição espacial dos elementos de controle é sempre adequada e não afeta a validade dos resultados, o que pode não ser verdade. A distribuição espacial de uma amostra pode determinar a validade de um processo de avaliação (ARIZA-LÓPEZ e ATKINSON-GORDO, 2008). Outro aspecto relevante não abordado neste trabalho foi o tamanho da área analisada. Espera-se que, quanto maior a área, maior o número de elementos de controle para cobri-la adequadamente, visando representar toda a área avaliada.

Estendendo a análise de literatura, observa-se que nem mesmo os trabalhos que apresentam os métodos de avaliação da acurácia posicional baseado em feições lineares sugerem como determinar o tamanho da amostra, sua distribuição espacial pela área avaliada ou como estes parâmetros podem influenciar os valores estimados. Os métodos *Buffer* Simples (GOODCHILD e HUNTER, 1997), *Buffer* Duplo (TVEITE e LANGAAS, 1999), Distância de Hausdorff (ABBAS, 1994) e Influência do Vértice (MOZAS-CALVACHE e ARIZA-LÓPEZ, 2011) são exemplos de métodos comumente utilizados e que não tratam do assunto.

Por fim, a falta de abordagem sobre o tamanho amostral e o padrão de distribuição espacial para o caso em que se utilize feições lineares no controle de qualidade cartográfica pode ser verificada ainda realizando-se uma análise sobre as normas de alguns países e padrões internacionais. Isso porque muitas agências nacionais de mapeamento utilizam os métodos baseados em pontos nos seus padrões para o controle posicional de produtos cartográficos (ARIZA-LÓPEZ e MOZAS-CALVACHE, 2012).

As normas da África do Sul para a produção da série de mapas nacionais são do ano de 2010, mas estas foram atualizadas nos últimos anos, sendo as versões mais recentes de 2017 a 2023. Essas normas, denominadas de Padrões para Séries de Ortofotomapas, são voltadas para a elaboração de ortofotomapas e apresentam os critérios para o controle posicional de mapas nas versões de papel e digital nas escalas de 1/10.000 (DRDLR, 2018a), 1/50.000 (DRDLR, 2023), 1/250.000 (DRDLR, 2017a) e 1/500.000 (DRDLR, 2017b).

Este conjunto de normas menciona somente a utilização de pontos bem definidos para a avaliação acurácia posicional, sem permitir a utilização de outros tipos de feições presentes nos mapas, como linhas. Esses documentos ainda não mencionam o tamanho que a amostra de controle deve possuir nem a sua distribuição espacial pela área analisada.

A norma utilizada atualmente na Austrália, a *Australian Map and Spatial Data Horizontal Accuracy Standard* (AMSDAS) de 2009, é baseada na *US National Standard for Spatial Data Accuracy* (NSSDA) apresentada por FGDC (1998) e compatível com a série ISO 19100 (ICSM, 2009). Segundo o documento, essa norma é aplicável a todos os tipos de feições, incluindo as geometrias de ponto, linha e polígono, não se limitando a pontos bem definidos, e para todos os tipos de dados,

sejam eles vetoriais ou matriciais (*raster*). Além disso, o texto estabelece que, no mínimo, 20 locais devem ser testados. Para avaliação utilizando feições lineares, esta norma recomenda aplicar *offsets* igualmente espaçados nas linhas testadas até sua interseção com a feição de referência, dada por uma fonte independente de melhor acurácia; similar ao que ocorre no *Buffer* Simples. Apesar da proposta ser interessante, considerando-se o fato de se adicionar outras primitivas geométricas na avaliação da acurácia posicional e de trazer um tamanho amostral, a norma não aborda ou deixa em aberto questões importantes. A primeira delas é com relação ao tamanho amostral. Apesar de trazer um valor mínimo, de 20 elementos, não é indicado o comprimento da amostra, dado que é essencial quando se utiliza linhas no controle de qualidade cartográfica. Tal crítica também é reforçada em Nero (2005). Outro fato tratado, mas que não é explicado, é sobre o cálculo da acurácia no caso em que não se utiliza pontos para o controle. Isso porque no capítulo em que apresenta os testes de acurácia, a norma menciona somente a comparação de coordenadas planimétricas a partir do RMSE (*Root Mean Square Error*). Não há uma métrica voltada para feições lineares ou de outros tipos, o que era esperado, tendo em vista que o documento propõe a sua utilização. Por fim, um tema não abordado é o padrão de distribuição espacial da amostra de controle.

Já o padrão adotado pelo México é composto pelo *Modelo de Calidad de Datos Espaciales* (INEGI, 2009) de 2009 e pela *Norma Técnica de Estándares de Exactitud Posicional* (INEGI, 2010) de 2010. Estes documentos tratam do controle posicional baseado somente em feições pontuais e não fazem menção ao tamanho da amostra de controle e nem a sua distribuição espacial pelo produto avaliado.

Portugal possui uma das normas mais recentes para o controle de qualidade cartográfica. Os documentos utilizados até então, as “Normas e Especificações Técnicas para a Produção de Cartografia e Ortofotocartografia” e os seus respectivos catálogos de objetos para as escalas 1/2.000 e 1/10.000, foram substituídos pelas “Normas e Especificações Técnicas para a Cartografia Topográfica Vetorial e de Imagem” em 2020. Essa nova especificação se enquadra nas iniciativas desenvolvidas pelo *Global Geospatial Information Management* das Nações Unidas, nas recomendações do *Open Geospatial Consortium* e nas normas ISO aplicáveis (DGT, 2020). Entretanto, apesar de ser uma norma recente e de trazer quase todos os elementos de qualidade da ISO, a especificação trata da avaliação da qualidade

posicional somente pela avaliação de pontos, sem indicar tamanho de amostra e o seu padrão de distribuição.

Outra norma recente é a utilizada na Espanha, a UNE 148002:2016 – *Control de la Calidad Posicional en Conjuntos de Datos Espaciales* (AENOR, 2016). Esta norma foi formulada para que a avaliação da qualidade posicional dos produtos cartográficos possa ser realizada com qualquer tipo de geometria: ponto, linha, polígono, volume, etc. A proposta de determinação do tamanho da amostra desse documento é a utilização da ISO 2859-1 (ISO, 1999) e da ISO 2859-2 (ISO, 2020). Diferentemente das normas anteriores, a UNE 148002:2016 além de afirmar que se deve assegurar que a amostra tenha uma boa distribuição espacial, estabelece critérios para garantir que isso ocorra e faz isso a partir de dois cenários: quando a área a ser avaliada é retangular, e quando ela é irregular. No caso de uma área retangular (como uma ortoimagem, por exemplo), a UNE 148002:2016 estabelece dois critérios a serem seguidos para garantir a boa distribuição espacial: deve-se dividir a área em 4 quadrantes, de modo que cada quadrante deva conter no mínimo 20% das amostras; e a distância entre as feições de checagem deve ser, aproximadamente, de um décimo da diagonal da área. Já para uma área irregular, esta norma afirma apenas que deve se ter cuidado para que a amostra cubra adequadamente o produto avaliado. Percebe-se que, apesar da boa iniciativa da norma de estabelecer estes critérios, pode haver problemas quando as feições de checagem não forem pontos. Não é simples dividir em quadrantes feições compridas, como estradas e rios, sem que uma feição atravesse para outro quadrante; e não é explicado como medir a distância entre feições que não sejam pontos para garantir que estas se distanciem de, aproximadamente, um décimo da diagonal da área. Além disso, como ocorre com a norma AMSDAS (ICSM, 2009), da Austrália, não é indicado o comprimento da amostra para o caso em que se utilize feições lineares para a avaliação posicional, o que é importante quando os elementos de controle são bidimensionais. Por fim, cabe destacar que essa norma não apresenta nenhuma métrica para se obter a discrepância posicional, nem mesmo a partir de pontos.

No Brasil, a qualidade dos produtos cartográficos é avaliada por meio do Decreto N° 89.817 de 1984 (BRASIL, 1984) e complementada pela Especificação Técnica Para Controle de Qualidade de Dados Geoespaciais (ET-CQDG) da Infraestrutura Nacional de Dados Espaciais (INDE) (DSG, 2016).

Enquanto o Decreto trata somente de pontos bem definidos, a ET-CQDG cita três tipos de amostragem espacial: amostragem de pontos de controle, amostragem de fenômenos contínuos e amostragem de feições geográficas (amostragem de objetos). Sendo que esta última engloba feições como rios, estradas, edifícios, etc. Entretanto, esta especificação técnica não traz nenhuma indicação ou métrica para se determinar a discrepância posicional a partir destes objetos, somente o fazendo para feições do tipo ponto. Além disso, não há neste documento regras ou indicação que propiciem uma boa distribuição espacial dos elementos de checagem.

A atual norma dos Estados Unidos da América (EUA), a *National Standard for Spatial Data Accuracy* (NSSDA) foi criada pelo *Federal Geographic Data Committee* (FGDC), que é o órgão dos EUA responsável por desenvolver ou adotar padrões relacionados a dados geoespaciais para a implementação da Infraestrutura Nacional de Dados Espaciais (*National Spatial Data Infrastructure - NSDI*) (FGDC, 1998). A NSSDA, em sua Parte 3 (*National Standard for Spatial Data Accuracy*), implementa uma metodologia estatística para estimar a acurácia posicional de mapas e dados geoespaciais digitais, baseada apenas em pontos bem definidos, sem a menção a outros tipos de elementos de controle. Esta parte da norma ainda estabelece uma amostra mínima de 20 pontos, que devem ser distribuídos para refletir a área de interesse. Em um anexo informativo, a norma fornece diretrizes para a distribuição espacial dos pontos de checagem. Neste anexo, o documento aconselha concentrar os pontos amostrais em áreas mais importantes e estabelece que para uma área retangular, em que se acredite ter acurácia posicional uniforme, os pontos podem ser distribuídos com um espaçamento de, aproximadamente, 10% do valor da diagonal, semelhante a um dos critérios da UNE 148002:2016 (AENOR, 2016).

Outra norma importante nos Estados Unidos é a ASPRS 2023, criada pela Sociedade Americana de Fotogrametria e Sensoriamento Remoto (*American Society of Photogrammetry and Remote Sensing*) (ASPRS, 2023). A norma cita que, dentre outros assuntos, metodologias para a avaliação da acurácia posicional utilizando feições lineares e a avaliação robusta da distribuição espacial e densidade das feições de checagem são necessárias. Entretanto, mesmo sendo um documento atualizado recentemente, o documento não aborda estes temas, e reconhece que estas são limitações presentes no documento.

Com relação às normas ISO, merece destaque a série ISO 19.100, que aborda as informações espaciais. Nesta série de documentos, aquela que se refere à qualidade dos dados é a ISO 19157:2013 (*Geographic information – Data quality*), que veio substituir as normas ISO 19138:2006 (*Geographic information – Data quality measures*), ISO 19114:2003 (*Geographic information – Quality evaluation procedures*) e ISO 19113:2002 (*Geographic information – Quality principles*) (ISO, 2013). Em seu anexo normativo, em que apresenta a lista de medidas padronizadas de qualidade de dados, este padrão descreve somente medidas baseadas em pontos, não mencionando a possibilidade de se utilizar feições lineares, tampouco aborda o tamanho da amostra de controle ou sua distribuição espacial pela área avaliada. Entretanto, vale mencionar que este padrão descreve os procedimentos gerais para avaliar a qualidade dos dados geográficos (ISO, 2013). Ou seja, ela apenas fornece uma base para que cada país crie suas próprias normas, não tendo o objetivo de ser uma norma regulamentadora. Por outro lado, vale notar que este padrão não menciona nem mesmo a possibilidade do uso de feições lineares para a avaliação da acurácia posicional, o que poderia servir de base e de incentivo para que as normativas nacionais tratassem destes procedimentos.

Outra importante norma internacional para a avaliação de produtos cartográficos é a STANAG 2215, criada pela Agência de Padronização da OTAN (Organização do Tratado do Atlântico Norte) (NSA, 2010). Assim como a ISO 19157, a STANAG 2215 é baseada na comparação de pontos homólogos. Apesar de admitir que alguns usuários podem necessitar de obter a acurácia posicional de feições lineares ao invés de pontos, este documento não aborda esta questão. A STANAG 2215 define um tamanho amostral fixo de 167 pontos, com a possibilidade do uso de um número menor de pontos de checagem. Com relação à distribuição espacial dos elementos de controle, o documento menciona somente que a amostra de checagem deve estar bem distribuída espacialmente sobre o produto avaliado.

Em resumo, de todas as normas analisadas, somente a da Austrália (AMSDAS) e a da Espanha (UNE 148002:2016) possibilitam se utilizar feições lineares para o controle posicional. E, dentre essas, somente a UNE 148002:2016 apresenta critérios para garantir a boa distribuição espacial da amostra. Ainda assim, há problemas quando esses critérios são utilizados em feições lineares.

Essa carência da abordagem do tamanho amostral de feições lineares na avaliação da acurácia posicional foi comprovada pela análise de artigos científicos, dos trabalhos que propõem os métodos e de normas e padrões internacionais. Baseado nessa análise de literatura, se desconhece metodologias, medidas ou ferramentas para a mensuração da distribuição espacial voltadas para feições do tipo linha. Esta lacuna na literatura representa um entrave para a utilização de feições lineares nos processos de controle da qualidade cartográfica.

### **3 Justificativa**

Uma das feições mais importantes em uma base de dados geográficos são as linhas (MOZAS-CALVACHE e ARIZA-LÓPEZ, 2018). Como mencionado, as linhas são primitivas geométricas mais complexas e com mais informações posicionais que os pontos (ângulos, direção, orientação, sinuosidade, comprimento, etc.), além de estarem presente em grande quantidade e, normalmente, com boa distribuição espacial nas bases cartográficas.

Diante disso, a discrepância posicional de uma linha bidimensional pode ser mais importante para Cartografia Digital e Sistemas de Informação Geográfica do que aquela do método tradicional, baseado em pontos (SHI, 1998). Além disso, trabalhos como os de Chrisman (1991) e Ariza-López e Mozas-Calvache (2012) defendem que não é possível supor que todas as feições de um produto cartográfico possam ter suas discrepâncias posicionais baseadas apenas em pontos bem definidos.

Entretanto, para se extrair todo o potencial da avaliação da acurácia posicional baseada em feições lineares, questões como a padronização do tamanho amostral e o seu padrão de distribuição espacial devem ser resolvidas, tendo em vista que tais parâmetros afetam os resultados e, conseqüentemente, o julgamento sobre os produtos avaliados.

## 4 Hipóteses

Diante do exposto, este trabalho apresenta as seguintes hipóteses:

- (i) É viável implementar uma metodologia para avaliar o padrão de distribuição espacial de feições lineares a partir dos seus vértices, baseado em uma modificação do método do Vizinho Mais Próximo para feições pontuais;
- (ii) É viável determinar, por meio de um processo de simulação, um tamanho amostral mínimo das feições lineares utilizadas no controle de qualidade cartográfica que seja representativo da população; e
- (iii) É possível implementar uma metodologia geral para a avaliação da acurácia posicional que utilize feições lineares que seja tão, ou mais, eficiente no controle posicional quanto o método tradicional por pontos.

## 5 Objetivos

### 5.1 Objetivo Geral

O objetivo principal desta pesquisa é desenvolver uma metodologia para a avaliação da qualidade posicional em um processo de controle de qualidade cartográfica utilizando feições lineares.

### 5.2 Objetivos Específicos

Considerando o objetivo geral, são propostos os seguintes objetivos específicos:

- Implementar uma metodologia para avaliar o padrão de distribuição espacial de feições lineares baseada na modificação do método do Vizinho Mais Próximo;
- Utilizar simulação computacional para gerar amostras, emulando diversas situações, como diferentes tamanhos amostrais e tamanhos da área de estudo, para definir qual tamanho de amostra mínimo é representativo da população, levando-se em consideração a distribuição espacial da amostra;
- Propor uma metodologia robusta, estatisticamente embasada, para a avaliação da acurácia posicional utilizando feições lineares; e

- Realizar a avaliação da acurácia posicional de alguns produtos cartográficos utilizando feições lineares segundo a metodologia proposta nesse trabalho. E realizar a mesma avaliação pelo método tradicional por pontos.

## 6 Estruturação do Trabalho

Este trabalho está dividido em cinco tópicos:

- Introdução geral – apresenta um panorama geral sobre o tema. Aborda o Controle de Qualidade Cartográfica, especialmente a avaliação da acurácia posicional utilizando feições lineares, a caracterização do problema, a justificativa, as hipóteses e os objetivos do estudo;
- Capítulo 1. Proposta de um método para a avaliação do padrão de distribuição espacial de feições lineares – neste capítulo é apresentado um método que permite inferir sobre o padrão de distribuição de linhas, baseado na modificação do método do Vizinheiro Mais Próximo para pontos;
- Capítulo 2. Determinação do tamanho amostral de feições lineares utilizadas no Controle de Qualidade Cartográfica por meio da Simulação Monte Carlo – utilizando um processo de simulação e o teste Kolmogorov-Smirnov, tem como objetivo determinar o tamanho amostral mínimo a ser utilizado na avaliação da acurácia posicional por meio de feições lineares, relacionando o resultado com o tamanho da área e considerando o padrão de distribuição espacial da amostra;
- Capítulo 3. Proposta de uma metodologia para avaliação da acurácia posicional utilizando feições lineares: o caso da norma brasileira e exemplos de aplicação – apresenta uma proposta geral para avaliação da acurácia posicional utilizando linhas, incluindo: detecção de *outliers*, determinação do tamanho amostral, avaliação do padrão de distribuição espacial da amostra, teste de tendência e teste de precisão; e
- Conclusões gerais – relata as conclusões obtidas com o desenvolvimento desta pesquisa.

## Referências Bibliográficas

ABBAS, I. **Base de Données Vectorielles et Erreur Cartographique. Problèmes Posés Par Le Contrôle Ponctuel. Une Méthode Alternative Fondée Sur La Distance de Hausdorff: Le Contrôle Linéaire.** Thèse de Doctorat en Informatique. Université Denis–Diderot, Paris, France, 1994.

AENOR - ASOCIACIÓN ESPAÑOLA DE NORMALIZACIÓN Y CERTIFICACIÓN. **UNE 148002:2016: Metodología de evaluación de la exactitud posicional de la información geográfica.** Madrid. 46p. 2016.

ANSI - AMERICAN NATIONAL STANDARDS INSTITUTE. **Spatial data transfer standard SDTS—Part 1: Logical specifications.** ANSI NCITS 320-1998, Washington, D.C, 1998.

ASPRS - AMERICAN SOCIETY FOR PHOTOGRAMMETRY AND REMOTE SENSING. ASPRS 2023: **Positional Accuracy Standards for Digital Geospatial Data.** Edition 2, Version 1.0. Disponível em < <https://www.asprs.org/revisions-to-the-asprs-positional-accuracy-standards-for-geospatial-data-2023> >. Acesso em 03 de mar. de 2024. 2023.

ARIZA-LÓPEZ, Francisco Javier; ATKINSON-GORDO, Alan David. Analysis of some positional accuracy assessment methodologies. **Journal of surveying Engineering**, v. 134, n. 2, p. 45-54, 2008.

ARIZA-LÓPEZ, F. J., MOZAS-CALVACHE, A. T., UREÑA-CÁMARA, M. A., ALBA-FERNÁNDEZ, V., GARCÍA-BALBOA, J. L., RODRÍGUEZ-AVI, J., RUIZ-LENDÍNEZ, J. J. Influence of sample size on line-based positional assessment methods for road data. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 66, n. 5, p. 708-719, 2011. DOI: 10.1016/j.isprsjprs.2011.06.003.

ARIZA-LÓPEZ, Francisco Javier; MOZAS-CALVACHE, Antonio Tomás. Comparison of four line-based positional assessment methods by means of synthetic data. **Geoinformatica**, v. 16, n. 2, p. 221-243, 2012.

ARIZA-LÓPEZ, F. J., GARCÍA-BALBOA, J. L., UREÑA-CÁMARA, M. A., REINOSO-GORDO, J. F. Propuesta de metodología para la evaluación de la calidad de elementos lineales 3D. In **X Congreso TOPCART - I Congreso Iberoamericano de Geomática Y. de la Tierra**, p. 16-19, 2012.

ARIZA-LÓPEZ, Francisco; RUIZ-LENDÍNEZ, Juan; UREÑA-CÁMARA, Manuel. Influence of Sample Size on Automatic Positional Accuracy Assessment Methods for Urban Areas. **ISPRS International Journal of Geo-Information**, v. 7, n. 6, p. 200, 2018.

ARIZA-LÓPEZ, Francisco Javier; RODRÍGUEZ-AVI, José; ALBA-FERNÁNDEZ, Virtudes. A Positional Quality Control Test Based on Proportions. In: **The Annual International Conference on Geographic Information Science**. Springer, Cham, p. 349-363, 2018.

ARIZA-LÓPEZ, F. J., GONZÁLEZ, P. B., PAU, J. M., TORRES, A. Z., PASCUAL, A. F. R., VERGARA, G. M., BALBOA, J. L. G. Geospatial data quality (ISO 19157-1): evolve or perish. **Revista Cartográfica**, (100), 129-154, 2020.

BLAKEMORE, M. Generalization and Error in Spatial Databases. **Cartographica**, v 21, p.131– 39, 1984.

BRASIL. Decreto N° 89.817 de 20 de junho de 1984. **Estabelece as Instruções Reguladoras das Normas Técnicas da Cartografia Nacional**. Brasil. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/decreto/1980-1989/D89817.htm](http://www.planalto.gov.br/ccivil_03/decreto/1980-1989/D89817.htm)>. Acesso em: 01 de mar. de 2023. 1984.

CHRISMAN, N. R. The error component in spatial data. **Geographical Information Systems**, v. 1, n. 12, p. 165-174, 1991.

CUENIN, René. Cartographie générale: Méthodes et techniques de production. **Eyrolles**, Vol. 1, 342p, 1972.

CUNHA, M. M.; SECATTO, G. Z.; GALINDO, J. R. F.; SANTOS, A. P. Proposta de um método de avaliação da acurácia posicional baseado na modificação do Buffer Simples. **Revista Brasileira de Cartografia**, v. 71, n. 4, p. 1193-1218, 2019.

DGT – DIREÇÃO GERAL DO TERRITÓRIO. **Normas e Especificações Técnicas para a Cartografia Topográfica Vetorial e de Imagem**. Portugal. Disponível em: <<https://www.dgterritorio.gov.pt/sites/default/files/ficheiros-cartografia/NormasEspecificacoesTecnicasCartTop.pdf>>. Acesso em: 01 de mar. de 2023. 2020.

DRDLR - DEPARTMENT RURAL DEVELOPMENT AND LAND REFORM. **Standard for the 1:10 000 Orthophoto Map Series**. Republic of South Africa. Disponível em:

<<http://www.ngi.gov.za/index.php/technical-information/standards-menu>>. Acesso em: 01 de mar. de 2023. 2018.

DRDLR - DEPARTMENT RURAL DEVELOPMENT AND LAND REFORM. **Standard for the 1:50 000 Orthophoto Map Series**. Republic of South Africa. Disponível em: <<http://www.ngi.gov.za/index.php/technical-information/standards-menu>>. Acesso em: 01 de mar. de 2023. 2023.

DRDLR - DEPARTMENT RURAL DEVELOPMENT AND LAND REFORM. **Standard for the 1:250 000 Orthophoto Map Series**. Republic of South Africa. Disponível em: <<http://www.ngi.gov.za/index.php/technical-information/standards-menu>>. Acesso em: 01 de mar. de 2023. 2017a.

DRDLR - DEPARTMENT RURAL DEVELOPMENT AND LAND REFORM. **Standard for the 1:500 000 Orthophoto Map Series**. Republic of South Africa. Disponível em: <<http://www.ngi.gov.za/index.php/technical-information/standards-menu>>. Acesso em: 01 de mar. de 2023. 2017b.

DSG - DIRETORIA DO SERVIÇO GEOGRÁFICO. **Especificação Técnica Para Controle de Qualidade de Dados Geospaciais (ET-CQDG)**. Brasília, 2016.

FGDC - FEDERAL GEOGRAPHIC DATA COMMITTEE. **Geospatial Positioning Accuracy Standards Part 3: National Standard for Spatial Data Accuracy**. United States. 1998.

FONSECA, I. G. R. **Avaliação da Acurácia Posicional Tridimensional Utilizando Feições Lineares**. Dissertação (Mestrado em Engenharia Civil). Viçosa: Universidade Federal de Viçosa. 84p. 2018.

GALO, M.; DAL POZ, A. P.; FERREIRA, F. M. O uso de feições no controle de qualidade em cartografia. In: XX CONGRESSO BRASILEIRO DE CARTOGRAFIA, **Anais.....**, Porto Alegre, RS, 2001.

GOODCHILD, M. F.; HUNTER, G. J. A simple positional accuracy measure for linear features, **International Journal of Geographical Information Science**, 11:3, 299-306, 1997, DOI: 10.1080/136588197242419.

HABIB, A.; MORGAN, M.; KIM, E. M.; CHENG, R. Linear Features in Photogrammetric Activities. In: **International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, 35:610–15. Istanbul, Turquia, 2004.

ICSM - INTERGOVERNMENTAL COMMITTEE ON SURVEY AND MAPPING. **Australian Map and Spatial Horizontal Data Accuracy Standard**. Australia. Disponível em: <<https://www.icsm.gov.au/publications/spatial-data-horizontal-accuracy-standard>>. Acesso em: 01 de mar. de 2023. 2009.

INEGI - INSTITUTO NACIONAL DE ESTADISTICA Y GEOGRAFIA. **Norma Técnica De Estandares De Exactitud Posicional**. Diario Oficial de la Federación el 23 de diciembre de 2010. México. 2010.

INEGI - INSTITUTO NACIONAL DE ESTADISTICA Y GEOGRAFIA. **Modelo de Calidad de Datos Espaciales: Anexo Técnico 2 - Evaluación de la Calidad de los Datos Espaciales**. Aguascalientes, México. 2009.

ISO - INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 2859-1:1999 - Sampling procedures for inspection by attributes — Part 1: Sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection**. Geneva, Switzerland, 87 p, 1999.

ISO - INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 19157: Geographic Information—Data Quality**. Geneva, Switzerland, 170 p, 2013.

ISO - INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 2859-2:2020 - Sampling procedures for inspection by attributes — Part 2: Sampling plans indexed by limiting quality (LQ) for isolated lot inspection**. Geneva, Switzerland, 46 p, 2020.

JAKOBSSON, A.; VAUGLIN, F. Report of a questionnaire on data quality in National Mapping Agencies. **CERCO Working Group on Quality. Comité Européen des Responsables de Cartographie Officielle**, Marne-la-Vallée, 2002.

KUBIK, Kurt. Relative and absolute orientation based on linear features. **ISPRS Journal of Photogrammetry and Remote Sensing**, 46(4): 199-204, 1991.

LA VEGA, P. G; ARIZA-LÓPEZ, F. J.; MOZAS-CALVACHE, A. T. Models for positional accuracy assessment of linear features: 2D and 3D cases. **Survey Review**, v. 48, n. 350, p. 347-360, 2016.

LUNARDI, O. A.; PENHA, A. L. T.; CERQUEIRA, R. W. O exército brasileiro e os padrões de dados geoespaciais para a INDE. In: **Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação**. 4, 2012, Recife. p. 001 – 008. 2012.

MOREIRA, J. M.; OLIVEIRA, G. A.; SILVA, P. L.; SANTOS, A. P.; MARQUES, E. T. Avaliação Da Qualidade Posicional Planimétrica De Ortofotos Usadas No Cadastro Territorial Multifinalitário Através De Feições Lineares. In: CONGRESSO DE CADASTRO MULTIFINALITÁRIO E GESTÃO TERRITORIAL, **Anais...**, Florianópolis, SC, 2018.

MOZAS-CALVACHE, A. T.; ARIZA-LÓPEZ, F. J. Methodology for positional quality control in cartography using linear features. **The Cartographic Journal**, v. 47, n. 4, p. 371-378, 2010.

MOZAS-CALVACHE, Antonio Tomás; ARIZA-LÓPEZ, Francisco Javier. Assessment of Displacements of Linestrings Based on Homologous Vertexes. **ISPRS International Journal of Geo-Information**, v. 7, n. 12, p. 473, 2018.

MOZAS-CALVACHE, A. T. **Control de calidad posicional en cartografía por elementos lineales**. Tese (Doutorado em Ingeniería Cartográfica, Geodésica y Fotogrametría). Jaén: Universidad de Jaén, Espanha. 249p. 2007.

MOZAS-CALVACHE, A. T.; ARIZA-LÓPEZ, F. J. New method for positional quality control in cartography based on lines. A comparative study of methodologies. **International Journal of Geographical Information Science**, v. 25, n. 10, p. 1681-1695. 2011. DOI:10.1080/13658816.2010.545063.

MOZAS-CALVACHE, Antônio Tomás. Positional quality assessment based on linear elements. **Revista cartográfica**, n. 103, p. 11-31, 2021.

NERO, Marcelo Antônio. **Propostas para o controle de qualidade de bases cartográficas com ênfase na componente posicional**. Tese (Doutorado em Engenharia). Universidade de São Paulo. 2005.

NERO, M. A.; CINTRA, J. P.; FERREIRA, G. F.; PEREIRA, T. A. J.; FARIA, T. S. A computational tool to evaluate the sample size in map positional accuracy. **Boletim de Ciências Geodésicas [online]**, vol.23, n.3, pp.445-460, 2017. <https://doi.org/10.1590/s1982-21702017000300030>.

NSA - NATO STANDARDIZATION AGENCY. **STANAG (Standardization Agreement) 2215: Evaluation of land maps, aeronautical charts and digital topographic data**. North Atlantic Treaty Organization (NATO). Bruxelas, 7ª Ed, 2010.

SANTOS, A. P., MEDEIROS, N. G., SANTOS, G. R., RODRIGUES, D. D. Controle de qualidade posicional em dados espaciais utilizando feições lineares. **Boletim de Ciências Geodésicas**, v. 21, n. 2, p. 233-250, 2015.

SANTOS, A. P. **Controle De Qualidade Cartográfica: Metodologias Para Avaliação Da Acurácia Posicional Em Dados Espaciais**. Tese (Doutorado em Engenharia Civil) – Universidade Federal de Viçosa, Viçosa, 188p, 2015.

SANTOS, A. P., MEDEIROS, N. G., SANTOS, G. R. D., RODRIGUES, D. D. Avaliação da acurácia posicional planimétrica em modelos digitais de superfície com o uso de feições lineares. **Boletim de Ciências Geodésicas**, v. 22, n. 1, p. 157-174. 2016.

SHI, W. A generic statistical approach for modelling error of geometric features in GIS. **International Journal Geographical Information Science**, 12(2), p. 131–143. 1998.

TONG, X.; SUN, T.; FAN, J.; GOODCHILD, M.F.; SHI, W. A statistical simulation model for positional error of line features in Geographic Information Systems (GIS). **Int. J. Appl. Earth Obs. Geoinf.** 21, p. 136–148, 2013.

TVEITE, Havard; LANGAAS, Sindre. An accuracy assessment method for geographical line data sets based on buffering, **International Journal of Geographical Information Science**, v. 13, n. 1, p. 27-47, 1999, DOI: 10.1080/136588199241445.

VAN NIEL, T. G.; MCVICAR, T. R. Assessing positional accuracy and its effects on rice crop area measurement: an application at Coleambally Irrigation Area. **Australian Journal of Experimental Agriculture**, v. 41, n. 4, p. 557-566, 2001.

YATES, Luke A.; BROOK, Barry W.; BUETTEL, Jessie C. Spatial pattern analysis of line-segment data in ecology. **Ecology**, v. 103, n. 2, p. 1-13, 2022.

# CAPÍTULO 1. PROPOSTA DE UM MÉTODO PARA A AVALIAÇÃO DO PADRÃO DE DISTRIBUIÇÃO ESPACIAL DE FEIÇÕES LINEARES

## Resumo

Usualmente, a avaliação da acurácia posicional de produtos cartográficos é realizada por meio das discrepâncias posicionais e técnicas baseadas em pontos. Entretanto, a utilização de feições lineares apresenta algumas vantagens sobre o método baseado em pontos. Dentre estas, pode-se mencionar que as feições lineares possuem mais informações geométricas e posicionais que os pontos e que, em uma base cartográfica, aproximadamente 80% das feições são linhas. Apesar destas vantagens, parâmetros importantes para a avaliação da acurácia utilizando linhas não foram estabelecidos ou determinados, por exemplo, o padrão de distribuição espacial. Mediante uma análise de literatura, diferentemente do que ocorre com os pontos, não se teve conhecimento de um método destinado à sua inferência em feições lineares. A distribuição espacial é um fator relevante, pois pode interferir nos resultados e determinar a validade de um processo de avaliação. Diante disso, este estudo propõe um método baseado na modificação do Método do Vizinho Mais Próximo para pontos, para avaliar o padrão de distribuição espacial de feições lineares. O método, em substituição à distância euclidiana tradicional, utilizada pelo método para pontos, propõe utilizar a Distância de Hausdorff como medida de espaçamento entre as linhas. O método proposto, denominado de Nearest Neighbor Method for Linear Features (NNMLF), foi aplicado a dados simulados e reais. Os resultados mostraram que o NNMLF foi eficaz em estimar o padrão de distribuição espacial esperado, até a terceira ordem, em todos os experimentos com dados simulados. Já a aplicação nos dados reais mostrou que o NNMLF é de simples aplicação. Além disso, este estudo apresenta uma breve discussão sobre como o número de feições pode influenciar no padrão de distribuição espacial de feições lineares. Por fim, o método proposto foi implementado em um *plugin* para QGIS para facilitar a sua utilização.

## 1 INTRODUÇÃO

Com o crescimento da disponibilidade das informações espaciais para os usuários, cresce também a necessidade de avaliar a qualidade destes dados e dos produtos gerados, para identificar se estes atendem aos requisitos dos usuários (XAVIER, ARIZA-LÓPEZ e UREÑA-CÁMARA, 2019). Nesse contexto, o Controle de Qualidade Cartográfica (CQC) avalia a qualidade de um produto cartográfico para que se possa ter confiabilidade nele, e possa ser utilizado de maneira adequada para atender às reais necessidades do usuário. Essa avaliação é realizada sobre os elementos de qualidade do produto, para os quais a norma ISO 19157 (ISO, 2013)

define que os elementos básicos da qualidade a serem observados e avaliados nos produtos cartográficos são: completude, consistência lógica, acurácia posicional, acurácia temporal, acurácia temática e usabilidade. Deve-se ressaltar que a acurácia posicional é o elemento mais utilizado para verificar a qualidade dos dados espaciais (JAKOBSSON e VAUGLIN, 2002; DROBNJAK et al., 2017).

Tradicionalmente, a avaliação da acurácia posicional é realizada por meio de técnicas baseadas em pontos (FGDC, 1998; VAN NIEL e MCVICAR, 2001; ARIZA-LÓPEZ et al., 2012; NERO et al., 2017; RUIZ-LENDÍNEZ, UREÑA-CÁMARA e ARIZA-LÓPEZ, 2017; ARIZA-LÓPEZ, RODRÍGUEZ-AVI e ALBA-FERNÁNDEZ, 2018; MOZAS-CALVACHE, 2021; NERO et al., 2022; SANTOS FILHO et al., 2022; ELIAS et al., 2023; TEIXEIRA e SCHMIDT, 2023; TRECANI et al., 2024). Isso ocorre através da comparação estatística das discrepâncias posicionais observadas entre pontos homólogos bem identificados nos produtos de teste e de referência.

Entretanto, o avanço tecnológico dos receptores GNSS (*Global Navigation Satellite System*) e o desenvolvimento de métodos de levantamento cinemático impulsionaram a possibilidade de aquisição e utilização de linhas no controle posicional (MOZAS-CALVACHE, 2007; MOZAS-CALVACHE, 2021). A utilização de linhas no CQC apresenta algumas vantagens em relação aos pontos. Por exemplo, Cuenin (1972), Thapa (1988) e Li (2006) afirmam que aproximadamente 80% das entidades de uma base cartográfica são constituídas de feições lineares. Já Mozas-Calvache e Ariza-López (2011) destacam que os elementos lineares possuem variadas informações geométricas e posicionais, apresentando vértices, ângulos, comprimento dos segmentos de reta, orientação, sinuosidade, entre outros. Estas informações adicionais podem auxiliar e proporcionar novas possibilidades no processo de avaliação da acurácia posicional. Logo, Mozas-Calvache e Ariza-López (2010) afirmaram que é possível que o controle da qualidade posicional em Cartografia por meio de feições lineares represente uma evolução dos métodos baseados em feições pontuais.

Entretanto, apesar destas vantagens, parâmetros importantes na avaliação da acurácia posicional utilizando feições lineares, como o tamanho amostral e o seu padrão de distribuição, ainda não foram estabelecidos ou determinados (ARIZA-LÓPEZ et al., 2011; ARIZA-LÓPEZ, RUIZ-LENDÍNEZ e UREÑA-CÁMARA, 2018).

A distribuição espacial de uma amostra, tema deste trabalho, pode determinar a validade e a qualidade de um processo de avaliação por amostragem. Isso porque uma má distribuição espacial afeta a representatividade da amostra. Isso significa que a amostra não retrata adequadamente a população, o que pode resultar em uma estimativa equivocada (ARIZA-LÓPEZ e ATKINSON-GORDO, 2008).

Quando se trata da distribuição espacial dos elementos de controle do tipo ponto existem métricas, estatísticas ou métodos que avaliam a sua distribuição, como o método do Vizinho Mais Próximo, apresentado por Clark e Evans (1954) e a Função K de Ripley, introduzida por Ripley (1977). No Controle de Qualidade Cartográfica, a aplicação do método do Vizinho Mais Próximo pode ser observada em trabalhos como Santos et al. (2016) e Silva (2020); enquanto a Função K de Ripley é utilizada nas pesquisas de Zanetti (2017), Oliveira et al. (2018) e Silva (2019), por exemplo. Além disso, normas de alguns países fornecem diretrizes para a distribuição espacial dos pontos de checagem, como a *National Standard for Spatial Data Accuracy* (NSSDA) (FGDC, 1998), dos Estados Unidos da América (EUA), e a UNE 148002:2016 – *Control de la Calidad Posicional en Conjuntos de Datos Espaciales* (AENOR, 2016), da Espanha.

Porém, quando a questão é o padrão de distribuição de feições lineares utilizadas na avaliação da acurácia posicional, o mesmo não ocorre, visto que o assunto parece ter sido pouco ou nada explorado em alguns aspectos. Essa questão pendente é comprovada por meio de uma análise na literatura, desconhecendo-se um método para avaliar a distribuição espacial para feições lineares.

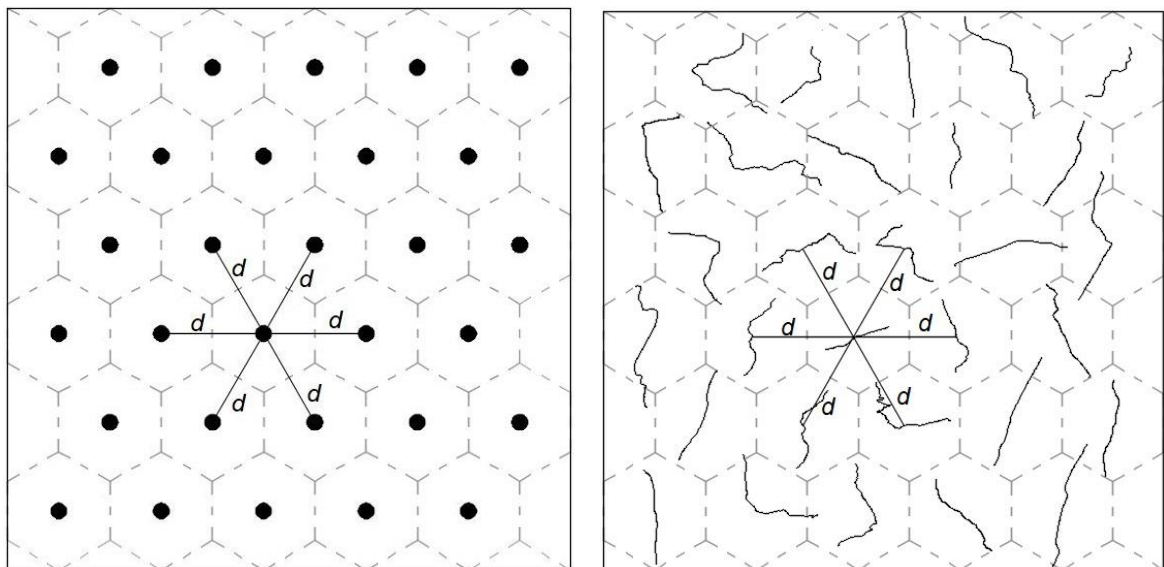
Diante disto, o objetivo deste trabalho foi propor um método de avaliação do padrão de distribuição espacial de feições lineares utilizadas na avaliação da acurácia posicional no CQC, o que seria uma abordagem inovadora neste assunto.

## **2 MÉTODO PROPOSTO**

O método proposto foi denominado Nearest Neighbor Method for Linear Features (NNMLF), ou Método do Vizinho mais Próximo para Feições Lineares (MVPFL), em português. Como o nome sugere, este método é baseado na modificação do Método do Vizinho Mais Próximo para pontos.

O Método do Vizinheiro Mais Próximo compara as distâncias médias observadas entre pontos com as distâncias esperadas em um conjunto teórico de feições pontuais que possui um padrão de distribuição aleatório. Esse conjunto teórico é obtido particionando-se a área de estudo em hexágonos de mesmo tamanho, com um ponto no centro de cada hexágono, fazendo com que cada elemento seja equidistante a outros seis e o espaço entre os elementos seja maximizado (CLARK e EVANS, 1954; LEE e WONG, 2001). A Figura 1a ilustra a distribuição teórica de 30 pontos, com a indicação das distâncias  $d$  entre seis deles.

Figura 1 – Padrão de distribuição espacial teórico para pontos (a) e linhas (b).



Fonte: (a) Adaptado de De Vos (1973), p. 309.

Este método tem distâncias como parâmetros de entrada, e essa é a premissa básica para que seja possível a adaptação do método tradicional para o caso de feições lineares. Sendo assim, a ideia principal do método proposto é obter um conjunto de distâncias que representem o espaçamento médio entre as feições lineares analisadas, e, como no método de referência, comparar com um padrão de distribuição espacial teórico. Este padrão teórico pode ser o mesmo, assumindo que a área é dividida em hexágonos de mesmo tamanho, com uma feição linear no centro de cada hexágono (Figura 1b). Isso porque, apesar de, diferentemente dos pontos, as linhas possuírem parâmetros geométricos, como direção, comprimento e sinuosidade, é possível perceber que as distâncias que maximizam o espaçamento são obtidas entre os centros dos hexágonos, assim como no caso dos pontos. Portanto, a

utilização do mesmo padrão teórico possibilita aplicar o arcabouço do método para pontos no NNMLF, cujos parâmetros de entrada também serão distâncias, mas entre linhas.

Como medida de espaçamento entre as linhas optou-se por utilizar a métrica da Distância de Hausdorff. A Distância de Hausdorff é amplamente utilizada para recuperação de informações e análise da similaridade geométrica entre objetos vetoriais (para pontos, linhas ou polígonos) ou imagens (HUTTENLOCHER et al., 1992; ARIZA-LÓPEZ e MOZAS-CALVACHE, 2012; CHEHREGHAN e ALI ABBASPOUR, 2017; MAROŠEVIĆ, 2018; WANG et al., 2019; Li et al., 2024; Liu et al., 2024). Segundo Atkinson-Gordo e Ariza-López (2002), entre as aplicações recorrentes desta distância estão a sua utilização na avaliação da qualidade posicional (MOZAS-CALVACHE e ARIZA-LÓPEZ, 2015; SANTOS et al., 2015; MOZAS-CALVACHE, PÉREZ-GARCÍA e FERNÁNDEZ-DEL CASTILLO, 2017; MOZAS-CALVACHE, UREÑA-CÁMARA e ARIZA-LÓPEZ, 2017; SAITO et al., 2019) e em processos de controle da eficácia na generalização cartográfica (ZHAI et al., 2017; GUO et al. 2019; LIU et al., 2020). Entretanto, para estas aplicações, a Distância de Hausdorff retrata a distância entre pares de linhas homólogas, que não é o caso desta pesquisa. Como dito, para o Nearest Neighbor Method for Linear Features, a ideia é utilizar a Distância de Hausdorff para representar o afastamento de feições lineares diferentes. Esta aplicação também é prevista por Hangouet (1995).

Sendo assim, o algoritmo do NNMLF é dado pela sequência de procedimentos listados a seguir.

**Passo 1:** Obter um conjunto de  $n$  linhas, o qual se deseja avaliar o padrão de distribuição espacial.

**Passo 2:** Obter o tamanho  $A$  da área da região em que se encontram as feições lineares a serem avaliadas.

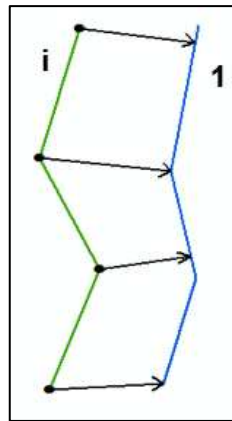
**Passo 3:** Extrair as coordenadas dos vértices de todas as linhas do conjunto de  $n$  linhas.

Como aborda Hangouet (1995), a Distância de Hausdorff não necessita ser calculada somente a partir de vértices. Entretanto, utilizar um ponto qualquer no segmento de uma linha para o cálculo tornaria a implementação do método proposto

mais difícil computacionalmente. Por isso, o método proposto se baseia somente nas coordenadas dos vértices das linhas.

**Passo 4:** Seja uma linha  $i$  pertencente ao conjunto de  $n$  linhas. Isto é,  $i \in \{1, 2, 3, \dots, n\}$ . Calcular a menor distância Euclidiana de cada vértice da linha  $i$  em relação a qualquer segmento da linha 1, como indicado na Figura 2. Neste passo, a distância Euclidiana é obtida entre um vértice da linha  $i$  e um vértice ou um ponto qualquer nos segmentos de reta que compõem a linha 1, escolhendo sempre o que fornece a menor distância. Repetir o processo, calculando a menor distância Euclidiana dos vértices da linha  $i$  para qualquer segmento das linhas 2, 3, ...,  $n$ , exceto para a própria linha  $i$ .

Figura 2 - Menores distâncias Euclidianas dos vértices da linha  $i$  para qualquer segmento da linha 1.



**Passo 5:** Repetir o **Passo 4** para todas as linhas pertencentes ao conjunto de  $n$  linhas. Isto é,  $\forall i \in \{1, 2, 3, \dots, n\}$ .

**Passo 6:** Para cada linha pertencente ao conjunto de  $n$  linhas, fazer a média de todas as menores distâncias calculadas com relação a todas as outras linhas.

**Passo 7:** Obter a Distância de Hausdorff entre todos os pares de linhas do conjunto de  $n$  linhas. Considere  $j$  uma linha também pertencente ao conjunto de  $n$  linhas e diferente de  $i$ , isto é,  $j \in \{1, 2, 3, \dots, n\}$  e  $i \neq j$ . Sendo  $d_{ij}$  a média das menores distâncias da linha  $i$  para a linha  $j$  e  $d_{ji}$  a média das menores distâncias da linha  $j$  para a linha  $i$ , a Distância de Hausdorff ( $dh$ ) é definida como o maior valor entre  $d_{ij}$  e  $d_{ji}$ , conforme a Equação 1.

$$dh = \max\{d_{ij}, d_{ji}\} \quad (1)$$

Como a média das menores distâncias da linha *i* para a linha *j* provavelmente será diferente da média das menores distâncias da linha *j* para a linha *i*, esta etapa é importante para se obter uma distância única entre cada par de linhas, como ocorre no caso de distância entre pontos. Segundo Mozas-Calvache (2007), esta diferença se dá pela assimetria deste tipo de medida, baseada em linhas. Hangouet (1995) cita a assimetria como uma das propriedades da Distância de Hausdorff.

Em sua abordagem clássica, os valores de  $d_{ij}$  e  $d_{ji}$  são dados pelas distâncias máximas das menores distâncias da linha *i* para a linha *j* e da linha *j* para a linha *i*, respectivamente. Entretanto, Maiseli (2021) alerta para o fato de que, por utilizar valores máximos no cálculo de  $d_{ij}$  e  $d_{ji}$ , esta forma de se obter a Distância de Hausdorff se torna muito sensível a erros grosseiros ou ruídos. Diante disso, para contornar este problema, Mozas-Calvache (2007) propõe utilizar a média das menores distâncias para se obter  $d_{ij}$  e  $d_{ji}$ , em substituição aos valores máximos. Por isso, esta maneira de calcular a Distância de Hausdorff, baseada na média, foi utilizada nesta pesquisa e no método NNMLF proposto.

**Passo 8:** Para cada linha, selecionar a menor Distância de Hausdorff observada, que representa a distância para o vizinho mais próximo ( $dv_i$ ).

Esta etapa deverá ser modificada quando se desejar o vizinho mais próximo de ordens maiores do que 1. Por exemplo, se deseja-se obter o vizinho mais próximo de segunda ordem, deve-se selecionar a segunda menor Distância de Hausdorff observada; e assim por diante.

A descrição do padrão de distribuição de um conjunto de feições baseado somente na distância ao vizinho mais próximo de primeira ordem não é completa, pois desconsidera as demais relações espaciais (CLARK e EVANS, 1954). Sendo assim, é interessante a utilização de ordens superiores na análise do padrão de distribuição espacial, independentemente da feição utilizada. Com os passos aqui apresentados é possível avaliar o NNMLF até a sexta ordem.

**Passo 9:** Com o conjunto de menores distâncias obtidas no passo anterior, como mostra Clark e Evans (1954) e Lee e Wong (2001), deve-se calcular o índice **R**, dado pela Equação 2.

$$R(k) = \frac{R_{OBS}(k)}{R_{ESP}(k)} \quad (2)$$

Sendo, que  $R_{OBS}(k)$  e  $R_{ESP}(k)$  são dados pelas Equações 3 e 4, respectivamente.

$$R_{OBS}(k) = \frac{\sum_{i=1}^n dv_i(k)}{n} \quad (3)$$

$$R_{ESP}(k) = \gamma_{1k} \sqrt{\frac{A}{n}} \quad (4)$$

Onde,

- $k$  é a ordem do vizinho mais próximo;
- $R_{OBS}(k)$  é a média observada das distâncias de cada linha para o seu  $k$  vizinho mais próximo;
- $R_{ESP}(k)$  é a média esperada das distâncias entre os  $k$  vizinhos mais próximos para uma distribuição aleatória;
- $dv_i(k)$  é a Distância de Hausdorff de uma linha  $i$  ao seu  $k$  vizinho mais próximo;
- $\gamma_{1k}$  é uma constante, dada em função da ordem  $k$ , como pode ser observado na Tabela 1;
- $n$  é o número de linhas cujo padrão de distribuição espacial está sendo avaliado;
- $A$  é a área da região em análise.

Tabela 1 - Constantes para o cálculo do vizinho mais próximo, em função da ordem.

| Ordem (k) | $\gamma_1$ | $\gamma_2$ |
|-----------|------------|------------|
| 1         | 0,5000     | 0,2613     |
| 2         | 0,7500     | 0,2722     |
| 3         | 0,9375     | 0,2757     |
| 4         | 1,0937     | 0,2775     |
| 5         | 1,2305     | 0,2784     |
| 6         | 1,3535     | 0,2789     |

Fonte: Wong e Lee (2005).

**Passo 10:** Aplicar o Teste Z. Clark e Evans (1954) afirmam que aplicar um teste de significância, para avaliar se a distância média observada para o vizinho mais próximo é estatisticamente igual a distância média esperada da distribuição aleatória, aumenta a confiabilidade do método. A estatística  $Z$  é dada pela Equação 5 (LEE e WONG, 2001).

$$Z(k) = \frac{R_{OBS}(k) - R_{ESP}(k)}{SE_r(k)} \quad (5)$$

Onde,  $\mathbf{SE}_r(\mathbf{k})$  é dado pela Equação 6.

$$SE_r(k) = \gamma_{2k} \sqrt{\frac{A}{n^2}} \quad (6)$$

Sendo,

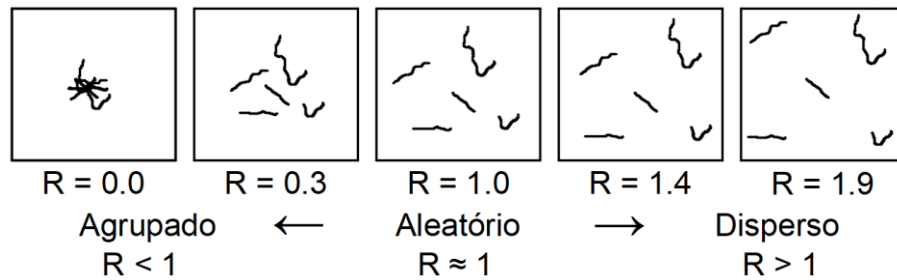
- $\gamma_{2k}$  é uma constante, dada em função da ordem  $\mathbf{k}$  (Tabela 1);
- $\mathbf{SE}_r(\mathbf{k})$  é o erro-padrão da diferença entre os valores médios das distâncias esperadas e das distâncias observadas entre os vizinhos mais próximos para a ordem  $\mathbf{k}$ . Descreve a probabilidade de que quaisquer diferenças ocorrerem puramente por acaso.

Neste teste, a hipótese nula é que os dados apresentam um padrão de distribuição espacial aleatório. Se o valor de  $\mathbf{Z}$  calculado para a ordem  $\mathbf{k}$  (Equação 5) for maior que o valor de  $\mathbf{Z}$  tabelado, rejeita-se a hipótese nula.

Se a hipótese nula não for rejeitada, significa que  $\mathbf{R}_{\text{OBS}}(\mathbf{k})$  é estatisticamente igual a  $\mathbf{R}_{\text{ESP}}(\mathbf{k})$ , tendo em vista que  $\mathbf{R}_{\text{ESP}}(\mathbf{k})$  representa o padrão aleatório esperado. Em outras palavras, significa dizer que  $\mathbf{R}(\mathbf{k})$  é estatisticamente igual a 1, já que ele é dado pela razão entre  $\mathbf{R}_{\text{OBS}}(\mathbf{k})$  e  $\mathbf{R}_{\text{ESP}}(\mathbf{k})$ . Caso a hipótese nula seja rejeitada e o valor de  $\mathbf{R}(\mathbf{k})$  for menor que 1, pode-se dizer que o valor de  $\mathbf{R}_{\text{OBS}}(\mathbf{k})$  é estatisticamente menor que o valor esperado para o padrão aleatório. Nesse caso, o padrão de distribuição espacial é agrupado. Por outro lado, se a hipótese nula for rejeitada e o valor de  $\mathbf{R}(\mathbf{k})$  for maior que 1, pode-se dizer que o valor de  $\mathbf{R}_{\text{OBS}}(\mathbf{k})$  é estatisticamente maior que o valor esperado para o padrão aleatório, sendo o padrão de distribuição espacial considerado disperso.

Uma representação gráfica dos valores do índice  $\mathbf{R}$  e sua relação com o padrão de distribuição espacial observado é apresentada Figura 3.

Figura 3 - Relação do índice R com o padrão de distribuição espacial de feições lineares.



Fonte: Adaptado de Santos et al. (2016).

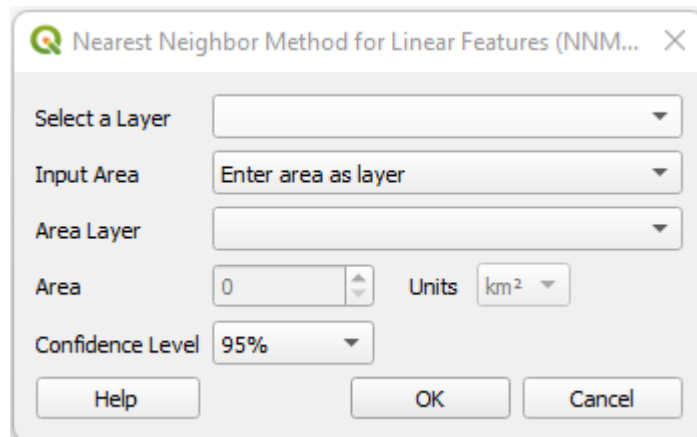
Para ilustrar a aplicação e avaliar o desempenho do método proposto, foram realizados três experimentos. Além disso, para facilitar a aplicação do NNMLF, foi implementada uma aplicação computacional baseada em sua metodologia.

Uma desvantagem do NNMLF é o seu custo computacional. Por extrair todos os vértices e calcular a Distância de Hausdorff destes pontos para todas as linhas, o método demanda um considerável esforço computacional. Apesar de não impedir a utilização do método, deve-se levar em consideração o tempo de processamento para conjuntos de dados que possuam linhas com uma grande densidade de pontos, como é comum em levantamentos com GNSS utilizando o método de posicionamento cinemático. Esta limitação pode ser contornada aplicando algum método de generalização de elementos lineares antes da execução do NNMLF, como o algoritmo Douglas-Peucker (DOUGLAS e PEUCKER, 1973).

### 3 IMPLEMENTAÇÃO DE UMA APLICAÇÃO COMPUTACIONAL PARA A UTILIZAÇÃO DO NNMLF

Para facilitar a utilização do NNMLF em um Sistema de Informações Geográficas (SIG) e visando a difusão do método proposto, foi elaborado um *plugin* para o QGIS versão 3.0.0 ou superior (QGIS, 2023). Este *plugin*, denominado de “Nearest Neighbor Method for Linear Features (NNMLF)”, é apresentado na Figura 4.

Figura 4 – Plugin Nearest Neighbor Method for Linear Features (NNMLF).

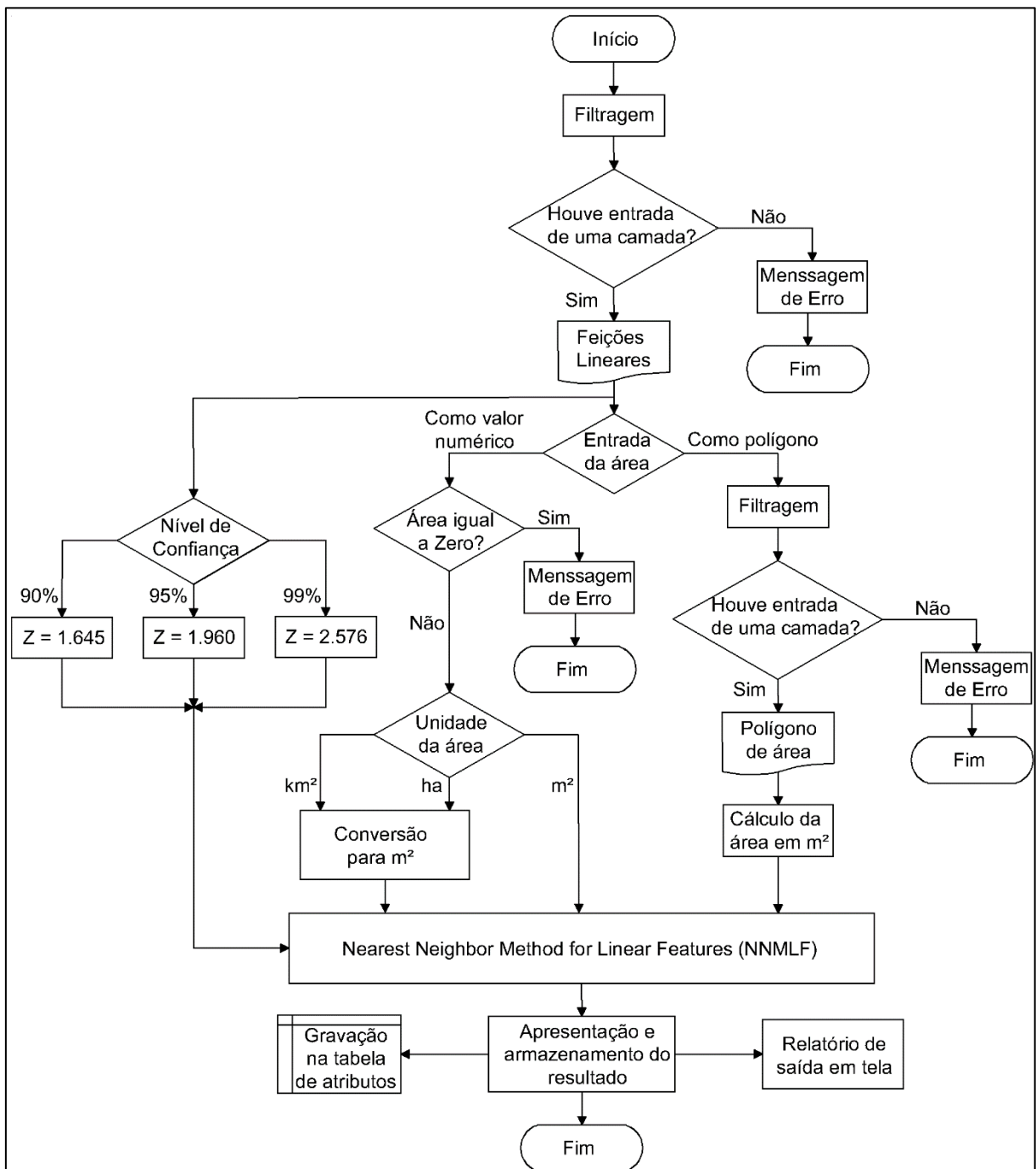


Esta aplicação computacional, produto da implementação da metodologia descrita na Seção 2, cujo código foi escrito em Python, tem como parâmetros de entrada: as feições lineares cujo padrão de distribuição espacial se deseja saber, a área e o nível de confiança desejado para o Teste Z. A Figura 5 traz o fluxograma de trabalho do *plugin*.

A primeira entrada solicitada é a camada com as feições lineares. A fim de evitar equívocos por parte do usuário, é realizada uma filtragem inicial e só é permitida a seleção de camadas vetoriais do tipo linha, contemplando todas as variações dos tipos `LineString` e `MultiLineString`. Além disso, na entrada dos dados, há uma condição que impede a execução do método caso nenhuma camada do tipo linha tenha sido selecionada, sendo exibida uma mensagem informando o erro.

Em seguida, o usuário deve fornecer a área. Isto pode ser feito de duas maneiras: fornecendo a área como valor numérico ou entrando com uma camada da área de estudo. Estas são opções mutuamente exclusivas, quando o usuário seleciona uma opção, a outra é desabilitada automaticamente, e vice-versa.

Caso o usuário opte por entrar com o valor numérico, é realizada uma verificação sobre este valor. A interface do *plugin* não permite valores negativos. Portanto, é necessário somente verificar se o valor da área é diferente de zero. Caso este valor seja zero, é exibida uma mensagem de erro avisando o usuário que o valor da área deve ser diferente de zero e, então, o processo é terminado. Caso contrário, ele deve selecionar as unidades da área. Há três opções: quilometro quadrado ( $\text{km}^2$ ), hectare (ha) e metro quadrado ( $\text{m}^2$ ).

Figura 5 - Fluxograma do *plugin NNMLF*.

Se o usuário optar por entrar com uma camada da área de estudo, também é realizada filtragem para que seja possível selecionar camadas do tipo polígono, abrangendo todas as variações dos tipos Polygon e MultiPolygon. Em seguida, é realizada a verificação se, ao selecionar a opção de adicionar uma camada, o usuário realmente a inseriu. Caso não haja nenhuma camada nesta entrada, é exibida uma mensagem de erro avisando o usuário que uma camada da área deve ser selecionada e, então, o processo é terminado. Caso contrário, a área do polígono é calculada.

Por fim, deve ser selecionado o nível de confiança desejado para o Teste Z. O *default* é 95%. Entretanto, há as opções de 90% e 99%. Além destas opções, o *plugin* dispõe também de um botão de ajuda, para fornecer ao usuário informações sobre o método NNMLF e sobre o próprio *plugin*.

Após a entrada dos dados necessários, o *plugin* calcula o Vizinho Mais Próximo para Feições Lineares até a Terceira Ordem, apresenta um relatório de saída em tela com todas as estatísticas e grava estes resultados na tabela de atributos da camada de linhas selecionada pelo usuário.

O *plugin* desenvolvido está disponível no repositório oficial de *plugins* em Python do QGIS, podendo ser instalado diretamente neste SIG. Caso deseje obter o código em Python, o usuário pode consegui-lo em: <[https://github.com/Marconi-Martins/NNMLF\\_QGIS.git](https://github.com/Marconi-Martins/NNMLF_QGIS.git)>. Além disso, os usuários interessados em utilizar este método contam com sua versão na linguagem R (R CORE TEAM, 2023). Esta versão está disponível em: <[https://github.com/Marconi-Martins/NNMLF\\_R.git](https://github.com/Marconi-Martins/NNMLF_R.git)>. Cabe destacar que esta última opção foi desenvolvida com programação paralela, visando uma melhor performance, e se encontra disponível no Apêndice A.

Vale ressaltar que o *plugin* NNMLF foi registrado no Instituto Nacional da Propriedade Industrial (INPI), com número de registro BR512023000701-3.

## 4 EXPERIMENTOS E RESULTADOS

Para analisar o comportamento e validar o método proposto, foram realizados três experimentos, com dados reais e simulados.

### 4.1 Experimento 1: Dados Simulados e as Ordens do NNMLF

No primeiro experimento foram criados dois conjuntos de dados simulados com padrões de distribuição espacial conhecidos para avaliar se o NNMLF é capaz de estimar estes padrões corretamente, até a Terceira Ordem.

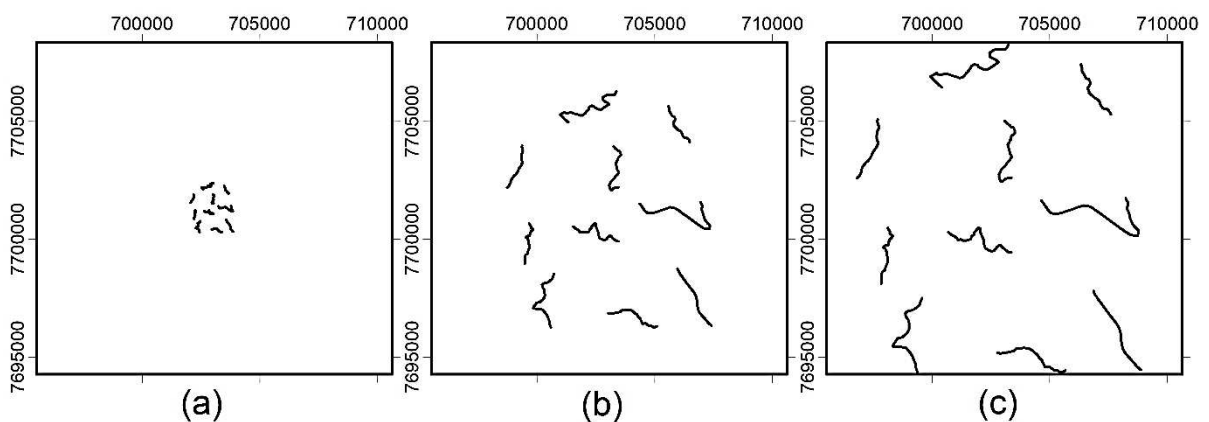
De todos os padrões de distribuição espacial possíveis (agrupado, aleatório e disperso), para qualquer ordem, o aleatório provavelmente é o mais difícil de simular. Isso porque os padrões agrupado e disperso podem ser emulados como casos

extremos, em que a distância entre as feições é muito pequena ou muito grande, respectivamente. Por outro lado, tentar retratar uma condição em que o padrão de distribuição espacial é aleatório pode ser especialmente complicado quando se trata de feições lineares. Isso porque estas, ao contrário dos pontos, apresentam comprimento, sinuosidade e direção, que influenciam nas distâncias e dificultam a previsão de um conjunto de distância entre as linhas que forneça o padrão aleatório. Somado a isso, como dito anteriormente, não se tem conhecimento de um método anterior para comparação e validação do desempenho do NNMLF.

Entretanto, pode-se utilizar uma estratégia para esta análise a partir da descrição de compressão de De Vos (1973). Este autor denomina de compressão o processo em que as distâncias entre as feições vão decrescendo enquanto a posição relativa entre elas permanece inalterada. Com a compressão, a distância para o vizinho mais próximo decresce, enquanto a distância esperada para um padrão de distribuição espacial aleatório permanece a mesma, tendo em vista que o número de feições e a área são os mesmos (DE VOS, 1973).

Então, para a avaliar se o NNMLF retorna corretamente o padrão de distribuição espacial esperado, foi utilizado o primeiro conjunto de linhas simuladas, dividido em três subconjuntos de dez linhas cada, cada um com um nível de compressão na mesma área, como pode ser visto na Figura 6.

Figura 6 - Primeiro conjunto de linhas simuladas. Padrões agrupado (a), aleatório (b) e disperso (c).



Para simular um grupo de linhas dispersas, as feições lineares foram dispostas de tal forma a ocupar toda a área disponível, maximizando a distância entre as feições,

como pode ser observado na Figura 6c. Já para retratar feições lineares agrupadas, foi realizada uma compressão nas linhas de modo que estas ocupassem uma região com dimensão de 2,25% da área, como mostra a Figura 6a. Por fim, para simular o padrão aleatório (Figura 6b), foi realizada uma compressão nas linhas para que estas ocupassem uma região equivalente à média das regiões dos padrões agrupado e disperso, para a mesma área. Desta forma, espera-se que o meio-termo entre os padrões agrupado e disperso forneça o padrão de distribuição espacial aleatório.

Sobre este conjunto de linhas foi aplicado o método proposto nesta pesquisa, o NNMLF. Esta etapa do experimento teve como objetivo avaliar somente o NNMLF de Primeira Ordem. Os resultados são apresentados na Tabela 2.

Tabela 2 - Resultado do NNMLF de Primeira Ordem para dados simulados.

|                             | Figura 6a | Figura 6b | Figura 6c |
|-----------------------------|-----------|-----------|-----------|
| <b>R</b> <sub>OBS</sub> (m) | 526,720   | 2514,775  | 3495,837  |
| <b>R</b> <sub>ESP</sub> (m) | 2281,149  | 2281,149  | 2281,149  |
| <b>R</b>                    | 0,231     | 1,102     | 1,532     |
| <b>Z</b>                    | -4,654    | 0,620     | 3,222     |
| <b>Z</b> tabelado           |           | 1,960     |           |
| <b>A</b> (km <sup>2</sup> ) |           | 208       |           |
| Nível de Confiança          |           | 95%       |           |
| Padrão Esperado             | Agrupado  | Aleatório | Disperso  |
| Resultado NNMLF             | Agrupado  | Aleatório | Disperso  |

Como pode ser visto na Tabela 2, os valores de **R** estão de acordo com o que era esperado para a estratégia utilizada. Isso mostra uma concordância entre a estratégia de se utilizar a compressão e o valor retornado pelo NNMLF. Com destaque para o índice **R** para o padrão aleatório, que ficou próximo de um.

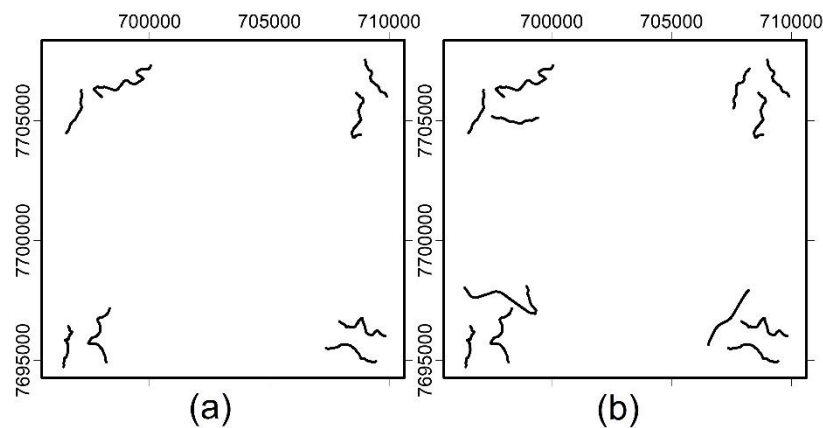
Para um Nível de Confiança de 95%, o valor de **Z** tabelado é de 1,96. Sendo assim, a hipótese nula do Teste Z, de que as feições lineares apresentam um padrão de distribuição espacial aleatório, foi rejeitada para o NNMLF de Primeira Ordem para os subconjuntos representados nas Figuras 6a e 6c, tendo em vista que em ambos os casos o valor de **Z** foi superior a 1,96 ( $|\mathbf{Z}(1)| > 1,96$ ). Como o valor de **R** do primeiro subconjunto é menor que um ( $\mathbf{R}(1) < 1$ ), então este apresenta um padrão de distribuição agrupado. Já o terceiro subconjunto apresenta um valor de **R** maior que

um ( $R(1) > 1$ ), resultando em um padrão disperso, segundo o NNMLF. Já para o segundo subconjunto, a hipótese nula não é rejeitada para o NNMLF de Primeira Ordem, pois  $|Z(1)| < 1,96$ , resultando em um padrão de distribuição espacial aleatório.

De maneira geral, os resultados mostraram que o método NNMLF de Primeira Ordem foi eficaz em estimar os padrões de distribuição espacial esperados para as linhas simuladas, obtendo sucesso nos três casos propostos.

Para avaliar o desempenho deste método nas ordens superiores (Segunda e Terceira Ordens), foi utilizado um segundo conjunto de linhas simuladas, como representado na Figura 7. A distribuição destas linhas foi inspirada nos trabalhos de Lee e Wong (2001) e Santos et al. (2016).

Figura 7 - Segundo conjunto de linhas simuladas. (a) Primeiro subconjunto e (b) segundo subconjunto.



O primeiro subconjunto (Figura 7a) é composto por oito feições lineares, dispostas de tal forma que a distância para o vizinho mais próximo de primeira ordem seja pequena e a distância para os vizinhos mais próximos de segunda e terceira ordens sejam grandes. Assim, para este subconjunto de dados simulados, espera-se que o NNMLF de Primeira Ordem seja agrupado, e os NNMLF de Segunda e Terceira Ordens sejam dispersos.

Já o segundo subconjunto (Figura 7b) é formado por 12 linhas. Estas feições foram simuladas de tal forma que as distâncias para os vizinhos mais próximos de primeira e segunda ordens sejam pequenas. Por outro lado, neste subconjunto a distância para o vizinho mais próximo de terceira ordem é muito grande. Sendo assim,

espera-se que os NNMLF de Primeira e Segunda Ordens sejam agrupados, e o NNMLF de Terceira Ordem seja disperso.

O método NNMLF foi aplicado a este conjunto de linhas e os resultados são apresentados na Tabela 3.

Tabela 3 - Resultado do NNMLF de Segunda e Terceira Ordens para dados simulados.

| Ordem                       | Primeiro subconjunto<br>(Figura 7a) |          |          | Segundo subconjunto<br>(Figura 7b) |          |          |
|-----------------------------|-------------------------------------|----------|----------|------------------------------------|----------|----------|
|                             | Primeira                            | Segunda  | Terceira | Primeira                           | Segunda  | Terceira |
| <b>R<sub>Obs</sub></b> (m)  | 1392,300                            | 9030,152 | 9914,311 | 1301,640                           | 1720,466 | 8277,309 |
| <b>R<sub>ESP</sub></b> (m)  | 2550,402                            | 3825,603 | 4782,004 | 2082,395                           | 3123,592 | 3904,490 |
| <b>R</b>                    | 0,546                               | 2,360    | 2,073    | 0,625                              | 0,551    | 2,120    |
| <b>Z</b>                    | -2,458                              | 10,602   | 10,322   | -2,485                             | -4,288   | 13,192   |
| <b>Z tabelado</b>           |                                     | 1,960    |          |                                    | 1,960    |          |
| <b>A</b> (km <sup>2</sup> ) |                                     | 208      |          |                                    | 208      |          |
| Nível de Confiança          |                                     | 95%      |          |                                    | 95%      |          |
| Padrão Esperado             | Agrupado                            | Disperso | Disperso | Agrupado                           | Agrupado | Disperso |
| Resultado NNMLF             | Agrupado                            | Disperso | Disperso | Agrupado                           | Agrupado | Disperso |

Os resultados mostraram que o método proposto retornou distâncias pequenas entre as feições para o NNMLF de Primeira Ordem do primeiro subconjunto de dados (Figura 7a) e para o NNMLF de Primeira e Segunda Ordens do segundo subconjunto (Figura 7b), assim como esperado. Da mesma forma, o método retornou valores grandes para as distâncias entre as linhas para o NNMLF de Segunda e Terceira Ordens do primeiro subconjunto (Figura 7a) e para o NNMLF de Terceira Ordem do segundo subconjunto de dados (Figura 7b).

Também considerando um Nível de Confiança de 95%, para o primeiro subconjunto (Figura 7a), a hipótese nula (padrão de distribuição espacial aleatório) foi rejeitada para as três ordens do NNMLF. Isso porque os valores de **Z** calculados foram superiores, em módulo, a 1,96:  $|\mathbf{Z}(1)| > 1,96$ ,  $|\mathbf{Z}(2)| > 1,96$  e  $|\mathbf{Z}(3)| > 1,96$ . Como  $\mathbf{R}(1) < 1$ , o NNMLF de Primeira Ordem é considerado agrupado. Por outro lado, como  $\mathbf{R}(2) > 1$  e  $\mathbf{R}(3) > 1$ , os NNMLF de Segunda e Terceira Ordens foram considerados dispersos. Para o segundo subconjunto (Figura 7b), a hipótese nula também foi rejeitada para as três ordens do NNMLF ( $|\mathbf{Z}(1)| > 1,96$ ,  $|\mathbf{Z}(2)| > 1,96$  e  $|\mathbf{Z}(3)| > 1,96$ ). Como  $\mathbf{R}(1) < 1$  e  $\mathbf{R}(2) < 1$ , os NNMLF de Primeira e Segunda Ordens foram

considerados agrupados. Já o NNMLF de Terceira Ordem resultou em um padrão disperso ( $R(3) > 1$ ).

Como consequência, assim como no caso do primeiro conjunto de dados, os resultados demonstraram que o método NNMLF foi eficaz em predizer os padrões de distribuição espacial esperado para as linhas simuladas, obtendo sucesso nos seis casos propostos.

De maneira geral, este primeiro experimento demonstrou que o Nearest Neighbor Method for Linear Features é eficaz em estimar todos os padrões de distribuição espacial possíveis (agrupado, aleatório e disperso) e em todas as três ordens (NNMLF de Primeira, Segunda e Terceira Ordens). Sendo assim, pode-se dizer que o método proposto tem bom desempenho e é efetivo para estimar o padrão de distribuição espacial em feições lineares.

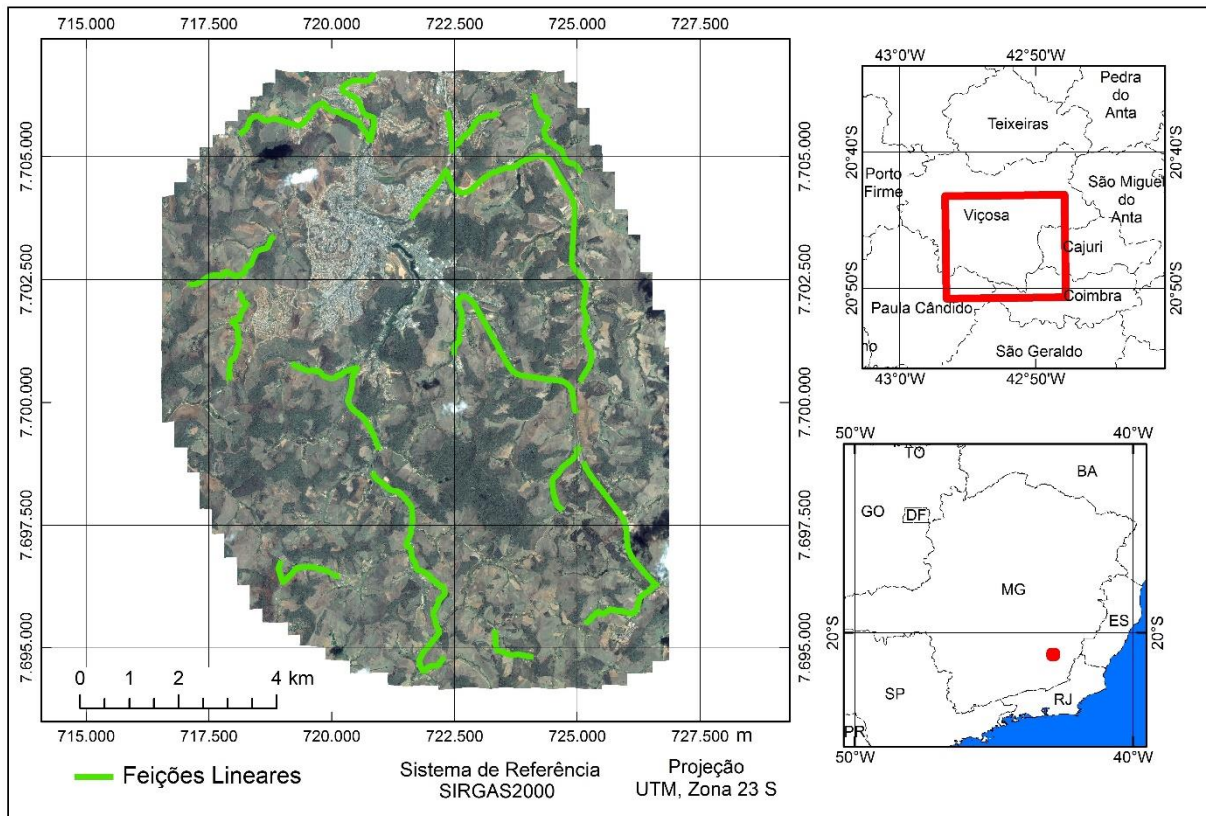
#### **4.2 Experimento 2: Dados Reais em um Estudo de Caso**

O segundo experimento proposto é um estudo de caso que teve por objetivo demonstrar o passo a passo da utilização do Método Vizinho Mais Próximo para Feições Lineares em uma situação real.

A área de estudo deste experimento (Figura 8) compreende parte do município de Viçosa, Minas Gerais, Brasil, e possui aproximadamente 110 km<sup>2</sup>. Nesta área foram utilizadas 14 estradas, coletadas por meio de receptores GNSS, que foram obtidas do trabalho de Santos et al. (2015).

Seguindo a sequência de procedimentos descritos na Seção 2, obtiveram-se as distâncias médias de cada uma das linhas com relação a todas as outras. O resultado, dado em metros, pode ser conferido na Tabela 4.

Figura 8 - Área de Estudo do estudo de caso do NNMLF.



Fonte: Adaptado de Cunha et al. (2019).

Tabela 4 - Distâncias médias entre as linhas, em metros.

| ID | 1        | 2        | 3       | 4       | 5        | 6       | 7       | 8       | 9       | 10       | 11      | 12       | 13       | 14      |
|----|----------|----------|---------|---------|----------|---------|---------|---------|---------|----------|---------|----------|----------|---------|
| 1  | 0        | 2339,22  | 4100,55 | 2568,92 | 2927,21  | 4230,00 | 4636,52 | 5227,44 | 8508,49 | 9260,03  | 8803,84 | 11555,08 | 11081,70 | 7460,88 |
| 2  | 1894,69  | 0        | 1502,96 | 896,72  | 4484,50  | 5700,09 | 3407,35 | 5366,95 | 6893,42 | 9442,90  | 7207,66 | 10073,62 | 10215,71 | 7267,44 |
| 3  | 3573,92  | 1130,69  | 0       | 652,48  | 6015,26  | 7116,67 | 3827,18 | 6257,20 | 6517,15 | 10064,48 | 6829,64 | 9562,64  | 10254,40 | 7875,32 |
| 4  | 3679,84  | 2288,36  | 1975,42 | 0       | 5217,21  | 6008,42 | 2183,53 | 4728,71 | 4730,33 | 8164,91  | 5038,60 | 7819,23  | 8302,70  | 6030,35 |
| 5  | 2679,44  | 4875,23  | 6705,11 | 3620,22 | 0        | 738,01  | 4488,11 | 2333,83 | 7719,49 | 6115,31  | 8064,51 | 10139,80 | 9065,93  | 5056,19 |
| 6  | 4056,35  | 5767,25  | 7449,63 | 4309,39 | 1170,92  | 0       | 4454,03 | 1359,50 | 7103,65 | 4715,32  | 7547,21 | 9244,59  | 7998,36  | 3990,73 |
| 7  | 5143,78  | 4302,24  | 4072,45 | 1479,24 | 5299,18  | 5407,74 | 0       | 3070,24 | 2479,08 | 5800,72  | 2775,02 | 5594,81  | 5728,14  | 3690,86 |
| 8  | 5117,57  | 5373,45  | 6516,92 | 3788,26 | 3127,63  | 2301,68 | 2412,46 | 0       | 4652,55 | 3681,19  | 5121,85 | 6860,32  | 5780,24  | 1829,95 |
| 9  | 7939,12  | 7073,03  | 6268,62 | 2041,73 | 7640,18  | 7075,41 | 1444,11 | 3780,72 | 0       | 4929,78  | 623,29  | 2773,51  | 3332,10  | 3090,60 |
| 10 | 8898,22  | 9132,99  | 9938,33 | 6898,97 | 6165,97  | 4192,93 | 5424,64 | 2974,26 | 5407,04 | 0        | 6197,68 | 5926,54  | 4143,03  | 2235,05 |
| 11 | 9334,73  | 8352,75  | 7215,35 | 3030,01 | 9145,94  | 8519,72 | 2485,23 | 5180,12 | 1319,11 | 5870,79  | 0       | 1433,43  | 3268,31  | 3894,71 |
| 12 | 10828,12 | 9951,05  | 8910,95 | 4705,50 | 10296,51 | 9277,63 | 4141,33 | 5923,62 | 2385,80 | 5804,76  | 816,36  | 0        | 2075,11  | 3638,04 |
| 13 | 10769,58 | 10246,82 | 9823,27 | 5648,88 | 9405,36  | 7961,44 | 5053,35 | 4881,59 | 3055,76 | 3799,33  | 3111,00 | 1645,20  | 0        | 1395,22 |
| 14 | 9036,70  | 8810,45  | 8972,21 | 5320,03 | 7228,94  | 5694,66 | 4546,70 | 2805,35 | 3404,44 | 2065,02  | 4135,88 | 3673,20  | 2044,09  | 0       |

O próximo passo foi obter a Distância de Hausdorff, que é definida como o maior valor entre as distâncias médias entre os pares de linhas. Como mencionado na descrição do método, as distâncias entre duas linhas não são iguais. Por exemplo, a distância média de todos os vértices da linha com ID 1 para qualquer segmento da linha com ID 2 é de 2339,22 m. Já a distância média de todos os vértices da linha com ID 2 para qualquer segmento da linha com ID 1 é de 1894,69 m. Sendo assim, para este caso, a Distância de Hausdorff entre as linhas ID 1 e 2, seria 2339,22 m.

Em seguida, deve-se obter o menor valor da Distância de Hausdorff para cada linha, com relação a todas as outras, que é o vizinho mais próximo. Fazendo isto até a terceira ordem, tem-se os valores apresentados na Tabela 5, em metros.

Tabela 5 - Vizinho Mais Próximo para Feições Lineares até Terceira Ordem.

| ID | Primeira Ordem (m) | Segunda Ordem (m) | Terceira Ordem (m) |
|----|--------------------|-------------------|--------------------|
| 1  | 2339,216           | 2927,211          | 3679,836           |
| 2  | 1502,956           | 2288,359          | 2339,216           |
| 3  | 1502,956           | 1975,417          | 4072,448           |
| 4  | 1975,417           | 2183,534          | 2288,359           |
| 5  | 1170,923           | 2927,211          | 3127,631           |
| 6  | 1170,923           | 2301,682          | 4229,999           |
| 7  | 2183,534           | 2479,078          | 2775,016           |
| 8  | 2301,682           | 2805,346          | 3070,235           |
| 9  | 1319,111           | 2479,078          | 2773,512           |
| 10 | 2235,050           | 3681,185          | 4143,029           |
| 11 | 1319,111           | 1433,431          | 2775,016           |
| 12 | 1433,431           | 2075,112          | 2773,512           |
| 13 | 2044,093           | 2075,112          | 3268,314           |
| 14 | 2044,093           | 2235,050          | 2805,346           |

Com este conjunto de menores distâncias, foram calculados os valores de distância média observada de cada linha para o seu  $k$  vizinho mais próximo ( $R_{OBS}$ ), a distância média esperada para o padrão de distribuição espacial aleatório ( $R_{ESP}$ ) para a ordem  $k$  e o índice  $R$ , com auxílio das Equações 3, 4 e 2, respectivamente. A Tabela 6 apresenta estes valores para as três primeiras ordens do NNMLF.

Tabela 6 - Valores de  $R_{OBS}$ ,  $R_{ESP}$  e Índice  $R$ 

|               | Primeira Ordem | Segunda Ordem | Terceira Ordem |
|---------------|----------------|---------------|----------------|
| $R_{OBS}$ (m) | 1753,036       | 2419,058      | 3151,534       |
| $R_{ESP}$ (m) | 1401,530       | 2102,295      | 2627,868       |
| $R$           | 1,251          | 1,151         | 1,199          |

A última etapa do método proposto consiste em aplicar o Teste Z para verificar se o índice  $R$  calculado é estatisticamente igual ao valor do  $R$  da distribuição aleatória. Os valores de  $Z$  para as três ordens são apresentados a seguir, na Tabela 7.

Tabela 7 - Valores de  $Z$ .

|     | Primeira Ordem | Segunda Ordem | Terceira Ordem |
|-----|----------------|---------------|----------------|
| $Z$ | 1,796          | 1,553         | 2,535          |

Considerando-se um Nível de Confiança de 95%, o valor de  $Z$  tabelado é de 1,96. Sendo assim, a hipótese nula do Teste Z, de que as feições lineares apresentam um padrão de distribuição espacial aleatório, é rejeitada para o NNMLF de Terceira Ordem ( $Z(3) > 1,96$ ) e não rejeitada para os o NNMLF de Primeira e Segunda Ordens ( $Z(1) < 1,96$  e  $Z(2) < 1,96$ ). Tendo em vista que  $R(3) > 1$ , tem-se os seguintes resultados:

- Nearest Neighbor Method for Linear Features de Primeira Ordem: Aleatório;
- Nearest Neighbor Method for Linear Features de Segunda Ordem: Aleatório;
- Nearest Neighbor Method for Linear Features de Terceira Ordem: Disperso.

Esta sequência de procedimentos demonstra que o método proposto nesta pesquisa é de simples aplicação, característica importante para o usuário final e essencial para implementação computacional.

### 4.3 Experimento 3: Dados Reais na Análise da Relação entre o Número de Feições Lineares e o Padrão de Distribuição Espacial

Quando se propõe um método novo, todas as variáveis que influenciam ou podem influenciar o resultado devem ser analisadas. O objetivo deste experimento foi avaliar se, para um determinado tamanho amostral, em termos de comprimento, diferentes números de feições lineares provocam resultados diferentes no padrão de distribuição espacial observado.

Esta análise é interessante, pois, como se pode observar pelas Equações 3 e 4, o número de feições influencia os valores de distâncias médias observada e esperada, o que pode influenciar no padrão de distribuição espacial. Entretanto, através de experimentos realizados, questões como a sinuosidade, orientação e comprimento podem interferir na Distância de Hausdorff observada entre as linhas e também interferir no resultado.

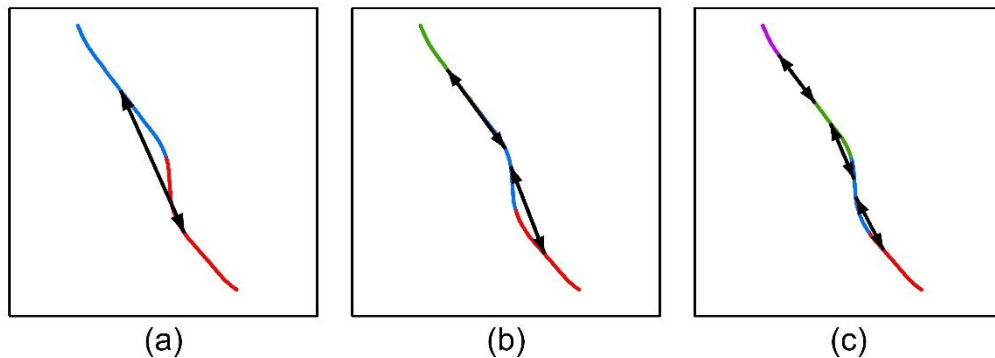
Sendo assim, para este experimento foram utilizadas as mesmas feições lineares (14 estradas) do experimento anterior (Figura 8). A partir destas linhas foram criados quatro conjuntos de dados. Além do conjunto original, com 14 estradas, para criar amostras com diferentes números de feições, cada linha foi dividida em duas, três ou quatro partes iguais. Isto resultou em quatro conjuntos de linhas, todos com o mesmo comprimento total, mas com 14, 28, 42 ou 56 linhas, dependendo da divisão. Sobre estes conjuntos de linhas foi aplicado o método Vizinho Mais Próximo para Feições Lineares (NNMLF). Os resultados podem ser encontrados na Tabela 8.

Tabela 8 - Influência do número de feições no padrão de distribuição espacial.

|                   |  | Número de feições           | 14        | 28        | 42       | 56       |
|-------------------|--|-----------------------------|-----------|-----------|----------|----------|
|                   |  | Comprim. amostra (m)        | 46249,040 |           |          |          |
|                   |  | Dividido por                | 1         | 2         | 3        | 4        |
|                   |  | <b>A</b> (km <sup>2</sup> ) | 110       |           |          |          |
|                   |  | Nível de Confiança          | 95%       |           |          |          |
| Primeira<br>Ordem |  | <b>R</b> <sub>OBS</sub> (m) | 1753,036  | 712,131   | 463,533  | 372,499  |
|                   |  | <b>R</b> <sub>ESP</sub> (m) | 1401,530  | 991,031   | 809,174  | 700,765  |
|                   |  | <b>R</b>                    | 1,251     | 0,719     | 0,573    | 0,532    |
|                   |  | <b>Z</b>                    | 1,796     | -2,850    | -5,297   | -6,708   |
|                   |  | Resultado NNMLF             | Aleatório | Agrupado  | Agrupado | Agrupado |
| Segunda<br>Ordem  |  | <b>R</b> <sub>OBS</sub> (m) | 2419,058  | 1435,980  | 796,224  | 592,794  |
|                   |  | <b>R</b> <sub>ESP</sub> (m) | 2102,295  | 1486,547  | 1213,760 | 1051,147 |
|                   |  | <b>R</b>                    | 1,151     | 0,966     | 0,656    | 0,564    |
|                   |  | <b>Z</b>                    | 1,553     | -0,496    | -6,143   | -8,991   |
|                   |  | Resultado NNMLF             | Aleatório | Aleatório | Agrupado | Agrupado |
| Terceira<br>Ordem |  | <b>R</b> <sub>OBS</sub> (m) | 3151,534  | 1890,348  | 1377,155 | 955,298  |
|                   |  | <b>R</b> <sub>ESP</sub> (m) | 2627,868  | 1858,184  | 1517,200 | 1313,934 |
|                   |  | <b>R</b>                    | 1,199     | 1,017     | 0,908    | 0,727    |
|                   |  | <b>Z</b>                    | 2,535     | 0,311     | -2,034   | -6,946   |
|                   |  | Resultado NNMLF             | Disperso  | Aleatório | Agrupado | Agrupado |

Como esperado, houve a diminuição das distâncias médias observada e esperada entre as feições ( $R_{OBS}$  e  $R_{ESP}$ , respectivamente) para as três ordens do NNMLF. A diminuição do valor de  $R_{OBS}$  se deu pela divisão das feições, o que fez com que a distância ao vizinho mais próximo fosse menor. A Figura 9 traz um exemplo para ilustrar como esta divisão afeta a Distância de Hausdorff, representada pelas setas. A Figura 9a representa uma feição dividida em duas partes iguais, a Figura 9b indica a divisão em três partes e a Figura 9c, em quatro partes.

Figura 9 - Exemplo de como o aumento do número de feições pode afetar a Distância de Hausdorff. Feição dividida em duas (a), três (b) e quatro (c) partes iguais.



Já a  $R_{ESP}$ , como pode ser constatado pela Equação 4, diminui a uma razão de  $\sqrt{n}$  com o aumento do número de feições. Isto pode ser observado na Tabela 8: quadruplicando-se o número de feições (de 14 para 56), a distância média esperada passou de 1401.530 m para 700.765 m para o NNMLF de Primeira Ordem, de 2102.295 m para 1051.147 m para o NNMLF de Segunda Ordem, e de 2627.868 m para 1313.934 m para o NNMLF de Terceira Ordem, em todos os casos ocorrendo uma razão de diminuição de  $\sqrt{4} = 2$ .

Já a diminuição do valor de  $R_{OBS}$  é mais complexa, pois depende de quanto a Distância de Hausdorff diminuiu, o que não é possível prever. Ainda que a Distância de Hausdorff permanecesse a mesma com o aumento do número de feições (o que não ocorre), o valor de  $R_{OBS}$  diminuiria a uma razão de  $n$  (Equação 3). Neste experimento, considerando 14 e 56 feições, a distância média esperada passou de 1753,036 m para 372,499 m (razão de diminuição de 4,7) para o NNMLF de Primeira Ordem, de 2419,058 m para 592,794 m (razão de diminuição de 4,1) para o NNMLF

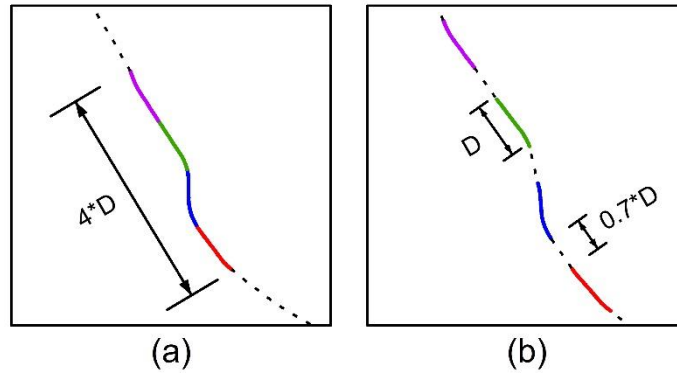
de Segunda Ordem, e de 3151,534 m para 955,298 m (razão de diminuição de 3,3) para o NNMLF de Terceira Ordem. Estes resultados e uma análise nas Equações 3 e 4 demonstram que, para uma amostra de mesmo comprimento, com o aumento do número de feições, o valor de  $R_{OBS}$  tende a diminuir a uma razão superior a  $R_{ESP}$ , fazendo com que o resultado do padrão de distribuição espacial observado seja agrupado, para as três ordens. Neste experimento, ao dobrar o número de feições (de 14 para 28), o NNMLF de Primeira Ordem passou de aleatório para agrupado. Já dividindo as 14 estradas originais em 3 partes iguais (número de feições igual a 42), todas as três ordens do NNMLF indicaram um padrão agrupado.

Apesar disso, se as linhas tivessem sido divididas e separadas, o resultado poderia ser diferente. Então, esta possibilidade também deve ser investigada. Por isso, a segunda parte deste experimento consistiu em realizar a divisão das linhas e aplicar uma distância entre as feições divididas, para observar o seu comportamento quando aplicado o Nearest Neighbor Method for Linear Features. Esta parte do experimento foi realizado para o subconjunto com 56 linhas.

Considerando uma feição linear do conjunto original com comprimento  $4 \cdot D$ , como indicado anteriormente, esta foi dividida em 4 partes iguais, cada uma de comprimento  $D$ . Para analisar a influência do aumento do número de feições no padrão de distribuição espacial, com a presença de espaçamento, foram empregadas as distâncias de  $0.1 \cdot D$ ,  $0.2 \cdot D$ ,  $0.3 \cdot D$ , ...,  $0.9 \cdot D$  e  $1.0 \cdot D$  entre as divisões, gerando-se 10 conjuntos de dados.

A Figura 10 apresenta o exemplo de uma linha dividida utilizada na primeira parte deste experimento (Figura 10a) e, em seguida, a mesma linha dividida, mas com um espaçamento entre as divisões (Figura 10b), que foi o caso analisado na segunda parte do experimento. É importante destacar que, ao introduzir distâncias entre as feições, estas sempre seguiram as estradas presentes na área de estudo (Figura 8), para retratar a realidade.

Figura 10 – Exemplo da aplicação de distância entre as linhas divididas. (a) Linha dividida sem espaçamento e (b) linha dividida e espaçada.



Os resultados desta análise podem ser observados na Tabela 9.



Estes resultados mostraram que o padrão de distribuição espacial observado foi agrupado para a segunda e terceira ordens do NNMLF, em todos os conjuntos de dados. Já para a primeira ordem, para distâncias de até 60% do comprimento das linhas, o padrão de distribuição indicado pelo NNMLF também foi o agrupado. Para distâncias iguais superiores a 70% do comprimento o padrão foi aleatório.

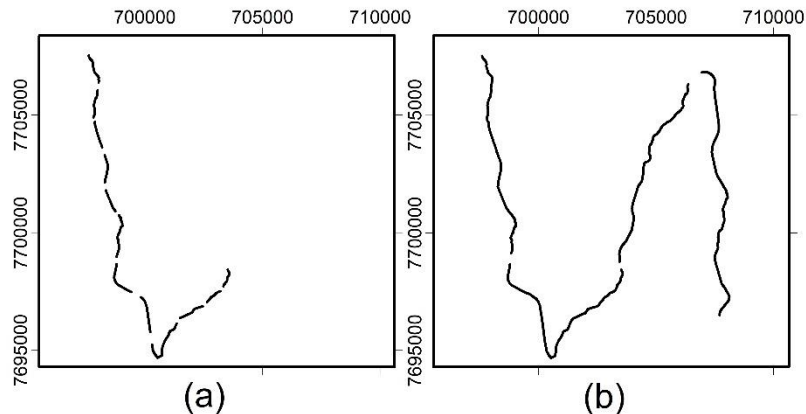
Cabe destacar que, segundo Clark e Evans (1954), a análise baseada somente na distância ao vizinho mais próximo de primeira ordem não é completa. Sendo assim, os conjuntos de dados que apresentaram um padrão de distribuição espacial aleatório para a primeira ordem do NNMLF, mas apresentaram um padrão agrupado para ordens superiores. Assim, poderiam ser considerados não adequados para uma análise completa das relações espaciais entre as feições.

Outro aspecto importante para se salientar nestes resultados é que 75% das feições (42 das 56) tiveram alteração de qual estrada foi considerada o vizinho mais próximo; sendo que, dentre essas mudanças, 57% ocorreram mais de uma vez. Este fato corrobora que a distância inserida entre as feições foi suficiente para mudar o cenário da primeira parte deste experimento.

Diante dos resultados deste experimento, pode-se dizer que, mesmo inserindo uma distância considerável entre as feições, o aumento do número de feições, para a mesma área e o mesmo tamanho amostral em termos de comprimento, tende a produzir padrões de distribuição espacial agrupados, comprometendo principalmente a análise de ordens superiores do NNMLF.

Em geral, o entendimento é que quanto maior o número de feições de checagem, melhor será a avaliação. Entretanto, este experimento demonstrou que o aumento do número de feições pode resultar em um padrão de distribuição espacial agrupado, o que não é recomendado para o Controle de Qualidade Cartográfica. Na realidade, pensando na avaliação da acurácia posicional utilizando feições lineares, deve-se preferir uma amostra que cubra melhor a área avaliada, independentemente do número de feições de checagem, desde que o padrão seja aleatório ou disperso, o que pode ser verificado com o método NNMLF. A Figura 11 traz um exemplo ilustrativo desta situação.

Figura 11 - Exemplo da relação entre o número de feições e o padrão de distribuição. (a) Doze linhas e (b) quatro linhas.



Neste exemplo, a Figura 11a é composta por 12 linhas, enquanto a Figura 11b apresenta quatro feições lineares. Entretanto, apesar de possuir um menor número de feições de checagem, a situação apresentada na Figura 11b possui um comprimento maior e representaria melhor as discrepâncias posicionais da área de interesse, por abranger melhor a área. Além disso, a Figura 11a apresenta um padrão de distribuição espacial agrupado, para as três ordens, enquanto a Figura 11b apresenta um padrão aleatório, também para as três ordens, sendo este o cenário recomendado.

Isso ocorre porque, diferentemente dos pontos, com as feições lineares é possível cobrir a área de interesse de forma satisfatória com um menor número de feições de checagem, desde que o tamanho amostral em termos de comprimento seja adequado. Esta característica das linhas se deve ao fato destas possuírem comprimento, o que não ocorre com os pontos.

Ainda sobre o impacto destes resultados no Controle de Qualidade Cartográfica, pode-se destacar que as 14 feições originais apresentaram padrão de distribuição espacial aleatório ou disperso (Tabela 8) para as três primeiras ordens, segundo o NNMLF. Quando estas feições foram divididas em quatro, na primeira parte do experimento, apresentaram um padrão agrupado para as três ordens. Segundo Santos et al. (2016), o padrão de distribuição espacial de uma amostra deve ser aleatório ou disperso, em um processo de avaliação da acurácia posicional. Portanto, segundo estes autores, as 14 linhas iniciais seriam adequadas para avaliar a qualidade posicional de um produto cartográfico. Entretanto, dividir as feições para se

obter um maior número de feições de checagem, visando, teoricamente, melhorar a avaliação da acurácia posicional, fez com que as mesmas feições lineares fossem consideradas inadequadas (Tabelas 8 e 9) para a avaliação da acurácia posicional.

Este fato reforça a importância deste experimento e os seus resultados: os dois conjuntos de linhas (14 e 56) possuem o mesmo comprimento, cobrem a mesma área e, provavelmente, apresentariam o mesmo resultado na avaliação da acurácia posicional. Entretanto, ao tratar o tamanho amostral em termos de número de feições de checagem, aumentando o número de feições, fez com que o conjunto de dados fosse impróprio para utilização no Controle de Qualidade Cartográfica.

Diante dos resultados deste experimento e desta discussão, para o controle posicional utilizando feições lineares, sugere-se tratar o tamanho amostral em termos de comprimento, independentemente do número de feições de checagem, desde que o padrão de distribuição espacial seja avaliado pelo método NNMLF e resulte em aleatório ou disperso. Isso evitaria problemas que o número de feições poderia causar e é mais adequado quando se trata de uma primitiva geométrica bidimensional.

## **5 CONCLUSÃO E CONSIDERAÇÕES FINAIS**

O desenvolvimento de um método para inferir sobre o padrão de distribuição espacial de linhas é de grande importância para a avaliação da acurácia posicional utilizando feições lineares no CQC. Além de permitir avançar em uma questão pouco explorada, como dito, a distribuição espacial de uma amostra pode determinar a validade de um processo de avaliação.

Durante a elaboração desta pesquisa, os resultados mostraram que o Nearest Neighbor Method for Linear Features (NNMLF) obteve êxito em estimar os padrões de distribuição espacial esperados em todos os experimentos propostos, com conjunto de dados simulados, para as três primeiras ordens do NNMLF. Já na aplicação do método proposto em dados reais, o NNMLF se mostrou de simples utilização. Por outro lado, uma desvantagem do método pode ser o tempo de processamento, quando este é utilizado em feições lineares com um grande número de vértices.

Uma discussão importante vinda dos resultados foi de como o número de feições pode impactar o padrão de distribuição espacial. Os resultados mostraram que aumentar o número de linhas, apenas visando ter mais feições de checagem, pode gerar um padrão de distribuição espacial agrupado, o que não é recomendado. Diante disso, recomenda-se utilizar o tamanho amostral em feições lineares em termos de comprimento, não como número de feições como ocorre com os pontos.

Um produto importante desta pesquisa é a ferramenta computacional criada para aplicação do NNMLF. O *plugin* desenvolvido será muito útil para os usuários que desejarem utilizar este método no QGIS.

Como recomendação para trabalhos futuros, para que se continue avançando em questões poucos exploradas na avaliação da acurácia posicional utilizando feições lineares, sugere-se o desenvolvimento de uma metodologia para a padronização do tamanho amostral.

Como conclusão geral, diante dos resultados apresentados, pode-se dizer que o NNMLF é um método robusto e pode ser utilizado na avaliação do padrão de distribuição espacial de feições lineares.

## REFERÊNCIAS BIBLIOGRÁFICAS

AENOR - ASOCIACIÓN ESPAÑOLA DE NORMALIZACIÓN Y CERTIFICACIÓN. **UNE 148002:2016: Metodología de evaluación de la exactitud posicional de la información geográfica**. Madrid. 46p. 2016.

ARIZA-LÓPEZ, Francisco Javier; ATKINSON-GORDO, Alan David. Analysis of some positional accuracy assessment methodologies. **Journal of surveying Engineering**, v. 134, n. 2, p. 45-54, 2008.

ARIZA-LÓPEZ, F. J., MOZAS-CALVACHE, A. T., UREÑA-CÁMARA, M. A., ALBA-FERNÁNDEZ, V., GARCÍA-BALBOA, J. L., RODRÍGUEZ-AVI, J., RUIZ-LENDÍNEZ, J. J. Influence of sample size on line-based positional assessment methods for road data. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 66, n. 5, p. 708-719, 2011. DOI: 10.1016/j.isprsjprs.2011.06.003.

ARIZA-LÓPEZ, F. J., GARCÍA-BALBOA, J. L., UREÑA-CÁMARA, M. A., REINOSO-GORDO, J. F. Propuesta de metodología para la evaluación de la calidad de

elementos lineales 3D. In **X Congreso TOPCART - I Congreso Iberoamericano de Geomática Y. de la Tierra**, p. 16-19, 2012.

ARIZA-LÓPEZ, Francisco Javier; MOZAS-CALVACHE, Antonio Tomás. Comparison of four line-based positional assessment methods by means of synthetic data. **Geoinformatica**, v. 16, n. 2, p. 221-243, 2012. DOI: 10.1007/s10707-011-0130-y.

ARIZA-LÓPEZ, Francisco Javier; RODRÍGUEZ-AVI, José; ALBA-FERNÁNDEZ, Virtudes. A Positional Quality Control Test Based on Proportions. In: **The Annual International Conference on Geographic Information Science**. Springer, Cham, p. 349-363, 2018.

ARIZA-LÓPEZ, Francisco; RUIZ-LENDÍNEZ, Juan; UREÑA-CÁMARA, Manuel. Influence of Sample Size on Automatic Positional Accuracy Assessment Methods for Urban Areas. **ISPRS International Journal of Geo-Information**, v. 7, n. 6, p. 200, 2018.

ATKINSON-GORDO, A. D.; ARIZA-LÓPEZ, F. J. Nuevo Enfoque Para El Análisis de La Calidad Posicional En Cartografía Mediante Estudios Basados En La Geometría Lineal. In: **XIV Congreso Internacional de Ingeniería Gráfica**. Espanha, 2002.

CHEHREGHAN, Alireza; ALI ABBASPOUR, Rahim. An assessment of the efficiency of spatial distances in linear object matching on multi-scale, multi-source maps. **International Journal of Image and Data Fusion**, v. 9, n. 2, p. 95-114, 2017. doi:10.1080/19479832.2017.1369175.

CLARK, P. J.; EVANS, F. C. Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. **Ecology**, 35(4), 445–453, 1954. DOI:10.2307/1931034

CUENIN, René. Cartographie générale: Méthodes et techniques de production. **Eyrolles**, Vol. 1, 342p, 1972.

CUNHA, M. M.; SECATTO, G. Z.; GALINDO, J. R. F.; SANTOS, A. P. Proposta de um método de avaliação da acurácia posicional baseado na modificação do Buffer Simples. **Revista Brasileira de Cartografia**, v. 71, n. 4, p. 1193-1218, 2019.

DE VOS, S. The use of nearest neighbor methods. **Tijdschrift voor economische en sociale geografie**, v. 64, n. 5, p. 307-319, 1973.

DOUGLAS, D. H.; PEUCKER, T. K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. **The Canadian Cartographer**, v. 10, n. 2, p. 112-122, 1973.

DROBNJAK, Siniša; BANKOVIĆ, Radoje; BAKRAČ, Saša; KOSTIĆ, Miodrag. Visualization of Horizontal Positional Accuracy Assessment Results for Digital Topographic Maps at Scale 1: 25000. In: **Sinteza 2017-International Scientific Conference on Information Technology and Data Related Research**. Singidunum University, p. 308-314, 2017.

ELIAS, E. N. N.; GIEHL, S.; AMORIM, F. R.; SCHMIDT, M. A. R.; CAMBOIM, S. P.; FERNANDES, V. O. QPEC: QGIS Toolkit for Evaluating Geospatial Data Positional Accuracy according to the Brazilian Cartographic Accuracy Standard. **Anuário do Instituto de Geociências**, v. 46., p. 1-16, 2023.

FGDC - FEDERAL GEOGRAPHIC DATA COMMITTEE. **Geospatial Positioning Accuracy Standards Part 3: National Standard for Spatial Data Accuracy**. United States. 1998.

GUO, Q.; XU, X.; WANG, Y.; LIU, J. Combined matching approach of road networks under different scales considering constraints of cartographic generalization. **IEEE Access**, v. 8, p. 944-956, 2019.

HANGOUE, Jean-Francois. Computation of the Hausdorff distance between plane vector polylines. In: **AUTOCARTO-CONFERENCE**. p. 1-10. 1995.

HUTTENLOCHER, D. P.; RUCKLIDGE, W. J.; KLANDERMAN, G. A. Comparing images using the Hausdorff distance under translation. In: **Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition**. IEEE, p. 654-656, 1992.

ISO - INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 19157: Geographic Information—Data Quality**. Geneva, Switzerland, 170 p, 2013.

JAKOBSSON, A.; VAUGLIN, F. Report of a questionnaire on data quality in National Mapping Agencies. **CERCO Working Group on Quality. Comité Européen des Responsables de Cartographie Officielle**, Marne-la-Vallée, 2002.

LEE, J.; WONG, D. W. S. **Statistical analysis with ArcView GIS**. John Wiley & Sons, 2001.

- LI, Zhilin. **Algorithmic foundation of multi-scale spatial representation**. CRC Press, 2006.
- LI, G.; ZHANG, X.; JIANG, L.; WANG, C.; HUANG, R.; LIU, Z. An approach for traffic pattern recognition integration of ship AIS data and port geospatial features. **Geospatial Information Science**, p. 1-28, 2024.
- LIU, B.; LIU, X.; LI, D.; SHI, Y.; FERNANDEZ, G.; WANG, Y. A vector line simplification algorithm based on the Douglas–Peucker algorithm, monotonic chains and dichotomy. **ISPRS International Journal of Geo-Information**, v. 9, n. 4, p. 251, 2020.
- LIU, J.; LI, J.; QIAO, L.; LI, M.; STEFANAKIS, E.; ZHAO, X.; HUANG, Q.; WANG, H.; ZHANG, C. QuadGridSIM: A quadrilateral grid-based method for high-performance and robust trajectory similarity analysis. **Transactions in GIS**, v. 28, n. 1, p. 83-107, 2024.
- MAISELI, Baraka Jacob. Hausdorff Distance with Outliers and Noise Resilience Capabilities. **SN Computer Science**, v. 2, n. 5, p. 1-12, 2021.
- MAROŠEVIĆ, Tomislav. The Hausdorff distance between some sets of points. **Mathematical Communications**, v. 23, n. 2, p. 247-257, 2018.
- MOZAS-CALVACHE, A. T. **Control de calidad posicional en cartografía por elementos lineales**. Tese (Doutorado em Engenharia Cartográfica, Geodésica y Fotogrametría). Jaén: Universidad de Jaén, Espanha. 249p. 2007.
- MOZAS-CALVACHE, A. T.; ARIZA-LÓPEZ, F. J. Methodology for positional quality control in cartography using linear features. **The Cartographic Journal**, v. 47, n. 4, p. 371-378, 2010.
- MOZAS-CALVACHE, A. T.; ARIZA-LÓPEZ, F. J. New method for positional quality control in cartography based on lines. A comparative study of methodologies. **International Journal of Geographical Information Science**, v. 25, n. 10, p. 1681-1695. 2011. DOI:10.1080/13658816.2010.545063.
- MOZAS-CALVACHE, A. T.; ARIZA-LÓPEZ, F. J. Adapting 2D positional control methodologies based on linear elements to 3D. **Survey Review**, v. 47, n. 342, p. 195-201, 2015.

MOZAS-CALVACHE, A. T.; PÉREZ-GARCÍA, J. L.; FERNÁNDEZ-DEL CASTILLO, T. Monitoring of landslide displacements using UAS and control methods based on lines. **Landslides**, v. 14, n. 6, p. 2115-2128, 2017.

MOZAS-CALVACHE, Antonio Tomás; UREÑA-CÁMARA, Manuel Antonio; ARIZA-LÓPEZ, Francisco Javier. Determination of 3D displacements of drainage networks extracted from digital elevation models (DEMs) using linear-based methods. **ISPRS International Journal of Geo-Information**, v. 6, n. 8, p. 234, 2017.

MOZAS-CALVACHE, Antônio Tomás. Positional quality assessment based on linear elements. **Revista cartográfica**, n. 103, p. 11-31, 2021.

NERO, M. A.; CINTRA, J. P.; FERREIRA, G. F.; PEREIRA, T. A. J.; FARIA, T. S. A computational tool to evaluate the sample size in map positional accuracy. **Boletim de Ciências Geodésicas [online]**, vol.23, n.3, pp.445-460, 2017. <https://doi.org/10.1590/s1982-21702017000300030>.

NERO, M. A.; PEREIRA, T. A. J.; RAMOS, M. P. Controle de qualidade posicional de ortoimagens a partir de imagens de satélite de boa resolução espacial de acordo com PEC-PCD. **Contemporânea - Revista de Ética e Filosofia Política**, v. 2, p. 1235-1251, 2022.

OLIVEIRA, G. D.; ALMEIDA, M. S.; MEDEIROS, N. G.; SANTOS, A. P.; DAL POZ, W. R. Correção geométrica de imagens orbitais a partir das coordenadas de vértices de imóveis certificados pelo INCRA. **Revista Brasileira de Cartografia**, v. 70, n. 1, p. 290-324, 2018.

QGIS. **QGIS Geographic Information System**. QGIS Association. Disponível em: <<http://www.qgis.org>>. Acessado em: 08 de fevereiro de 2023. 2023.

R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<https://www.R-project.org/>>. Acessado em: 08 de fevereiro de 2023. 2023.

RIPLEY, Brian David. Modelling spatial patterns. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 39, n. 2, p. 172-192, 1977.

RUIZ-LENDÍNEZ, Juan J.; UREÑA-CÁMARA, Manuel A.; ARIZA-LÓPEZ, Francisco J. A polygon and point-based approach to matching geospatial features. **ISPRS International Journal of Geo-Information**, v. 6, n. 12, p. 399, 2017.

SANTOS, A. P., MEDEIROS, N. G., SANTOS, G. R., RODRIGUES, D. D. Controle de qualidade posicional em dados espaciais utilizando feições lineares. **Boletim de Ciências Geodésicas**, v. 21, n. 2, p. 233-250, 2015.

SANTOS, A. P.; RODRIGUES, D. D.; SANTOS, N. T.; GRIPP JUNIOR, J. Avaliação Da Acurácia Posicional Em Dados Espaciais Utilizando Técnicas De Estatística Espacial: Proposta De Método E Exemplo Utilizando A Norma Brasileira. **Boletim de Ciências Geodésicas**, v. 22, n. 4, p. 630-650, 2016.

SANTOS FILHO, H.; CORNERO, C.; PEREIRA, A.; NERO, M. A. Cartographic Accuracy Standard (CAS) of the digital terrain model of the digital and continuous cartographic base of the state of Amapá: case study in the city of Macapá. **Boletim de Ciências Geodésicas**, v. 28, p. 1-20, 2022.

SAITO, Y. K.; VIANA, L. J. F.; SILVA, S. R.; FONSECA, I. G. R.; SANTOS, A. P. Influência da densidade de vértices nos métodos Distância de Hausdorff e Influência do Vértice. **Revista Brasileira de Cartografia**, v. 71, n. 2, p. 598-618, 2019.

SILVA, Rodrigo Pinto. **Análise da qualidade posicional de produtos gerados por LST e ARP em diferentes alturas de voo**. Dissertação (mestrado em Engenharia Civil) - Universidade Federal de Viçosa. Viçosa, 2019.

SILVA, Daniel da Costa. **Estudo da qualidade posicional e análise de padrões espaciais na distribuição de erros altimétricos em modelos digitais de elevação**. Dissertação (Mestrado em Ciências Geodésicas e Tecnologias da Geoinformação). Universidade Federal de Pernambuco. Recife, 2020.

TEIXEIRA, Moisés de Souza; SCHMIDT, Marcio Augusto Reolon. Avaliação da Acurácia Posicional de Dados Colaborativos do OpenStreetMap: Eixos Viários de Bairro no Município de Uberlândia (MG). **Revista Geociências**, v. 42, n. 2, p. 281-290, 2023.

THAPA, K. Automatic line generalization using zero-crossings. **Photogrammetric Engineering and Remote Sensing**, v. 54, n. 4, p. 511-517, 1988.

TRECCANI, D.; ADAMI, A.; BRUNELLI, V.; FREGONESE, L. Mobile mapping system for historic built heritage and GIS integration: a challenging case study. **Applied Geomatics**, p. 1-20, 2024.

VAN NIEL, T. G.; MCVICAR, T. R. Assessing positional accuracy and its effects on rice crop area measurement: an application at Coleambally Irrigation Area. **Australian Journal of Experimental Agriculture**, v. 41, n. 4, p. 557-566, 2001.

WANG, S.; GUO, Q.; XU, X.; XIE, Y. A study on a matching algorithm for urban underground pipelines. **ISPRS International Journal of Geo-Information**, v. 8, n. 8, p. 352, 2019.

WONG, D. W. S; LEE, J. **Statistical analysis of geographic information with ArcView GIS and ArcGIS**. John Wiley & Sons, 2005.

XAVIER, E.; ARIZA-LÓPEZ, F. J.; UREÑA-CÁMARA, M. A. Automatic evaluation of geospatial data quality using web services. **Revista cartográfica**, n. 98, p. 59-73, 2019.

ZANETTI, Juliette. **Influência do número e distribuição de pontos de controle em ortofotos geradas a partir de um levantamento por VANT**. Dissertação (mestrado em Engenharia Civil) - Universidade Federal de Viçosa. Viçosa, 2017.

ZHAI, J.; LI, Z.; WU, F.; XIE, H.; ZOU, B. Quality Assessment Method for Linear Feature Simplification Based on Multi-Scale Spatial Uncertainty. **ISPRS International Journal of Geo-Information**, v. 6, n. 6, p. 184, 2017.

## **CAPÍTULO 2. DETERMINAÇÃO DO TAMANHO AMOSTRAL DE FEIÇÕES LINEARES UTILIZADAS NO CONTROLE DE QUALIDADE CARTOGRÁFICA POR MEIO DA SIMULAÇÃO MONTE CARLO**

### **Resumo**

Em um processo de avaliação baseado em amostragem, o tamanho da amostra é sempre um tema muito discutido. Isso ocorre em qualquer área do conhecimento que utilize amostras, como o controle de qualidade cartográfica. Quando se trata da avaliação da acurácia posicional utilizando pontos, existem diversos estudos e normas que determinam qual o tamanho de amostra se deve utilizar. Entretanto, para o controle posicional utilizando linhas, não se tem conhecimento de estudo, norma ou documento que indique o tamanho da amostra que deve ser utilizado. Durante a realização desta pesquisa, se teve conhecimento somente de um trabalho (ARIZA-LÓPEZ et al., 2011) que discute o impacto do tamanho amostral de feições lineares na variabilidade da discrepância posicional, mas não há indicação do tamanho da amostra. Diante disso, este estudo propôs um método para determinar este tamanho amostral baseado na Simulação Monte Carlo. Para isso, foi utilizada uma base de dados composta por 46460 km de estradas de 10 províncias no leste da Espanha, e analisados os tamanhos de área de 1; 2,5; 5; 7,5; 10; 15; 20; 25; 50; 75; 100; 125; 150; 175; 200; 225 e 250 km<sup>2</sup>. Por fim, a comparação entre as discrepâncias fornecidas pelas amostras e pela população foi realizada com o teste Kolmogorov-Smirnov. Desta forma, foi possível realizar uma proposta do tamanho amostral de feições lineares utilizadas na avaliação da acurácia posicional, levando em consideração o padrão de distribuição espacial da amostra e o tamanho da área avaliada. Esta proposta foi realizada em forma de equação, que fornece o tamanho amostral com base no tamanho da área avaliada.

### **1 INTRODUÇÃO**

O tamanho da amostra é um dos aspectos principais de qualquer método de estimativa baseada em amostragem (ARIZA-LÓPEZ e ATKINSON-GORDO, 2008). E não é diferente no controle de qualidade cartográfica. No processo de avaliação da acurácia posicional, a determinação do tamanho amostra de controle é um dos temas mais discutidos e polêmicos, seja da perspectiva estatística ou mesmo prática. Isso porque a amostra de controle deve possuir um número de elementos suficientemente grande para garantir, sob determinado nível de confiança, a representatividade da população. Por outro lado, há a questão econômica, sendo que o número de elementos da amostra de controle deve ser o menor possível para minimizar os custos

do processo de avaliação da acurácia posicional (ARIZA-LÓPEZ e ATKINSON-GORDO, 2008; MOZAS-CALVACHE e ARIZA-LÓPEZ, 2010; ARIZA-LÓPEZ et al., 2011).

Para a avaliação da acurácia posicional baseada em feições pontuais, o tamanho amostral foi abordado em alguns trabalhos, como Nogueira Júnior (2003), Nogueira Júnior, Monico e Tachibana (2004), Nero (2005), Ariza-López et al. (2007), Ariza-López, Atkinson-Gordo e Nero (2007), Pereira e Nero (2012), Cintra e Nero (2015), Nero et al. (2017), entre outros.

Nogueira Júnior (2003) e Nogueira Júnior, Monico e Tachibana (2004) propõem uma equação para determinação do tamanho amostral. Esta equação é baseada na estatística Z para distribuição normal em função do intervalo de confiança, e utiliza o conceito de erro máximo permissível. Para a determinação deste erro, estes autores utilizou as tolerâncias de Brasil (1984), permitindo associar o tamanho da amostra à escala do produto cartográfico avaliado. Já Nero (2005) e Cintra e Nero (2015) propõem utilizar a norma ISO 2859-1 como referência para determinar o tamanho da amostra. Essa determinação é realizada a partir do tamanho da população, que é definida como o número de pontos existentes no mapeamento a ser avaliado quanto a sua qualidade posicional. Enquanto isso, Ariza-López et al. (2007) e Ariza-López, Atkinson-Gordo e Nero (2007) fazem uma revisão e análise crítica sobre temas importantes abordados em normas para avaliação posicional de produtos cartográficos da América Espanhola, incluindo o tamanho da amostra recomendado. Pereira e Nero (2012) realizaram uma análise comparativa sobre o tamanho da amostra e outros aspectos apresentados nas normas de diversos países, como África do Sul, México, Austrália, Portugal, Japão e Brasil. Por fim, Nero et al. (2017) apresentam uma metodologia para definir o tamanho da amostra em função do risco que o produtor ou o usuário pode ou quer assumir.

Além disso, em vários países, as normas e leis referentes ao assunto têm suas metodologias fundamentadas na comparação entre pontos homólogos (SANTOS et al., 2015). No Brasil, dentre manuais, normas e leis que abordam o tamanho amostral no controle de qualidade de cartográfica, destacam-se a Especificação Técnica Para Controle de Qualidade de Dados Geoespaciais (ET-CQDG) (DSG, 2016) e o Manual de Avaliação da Qualidade de Dados Geoespaciais (IBGE, 2019). Também merecem destaques as normas ASPRS (*American Society of Photogrammetry and Remote*

*Sensing*) 2014 (ASPRS, 2015) e a *National Standard for Spatial Data Accuracy* (NSSDA) (FGDC, 1998), dos Estados Unidos, a *Australian Map and Spatial Data Horizontal Accuracy Standard* (AMSDAS) da Austrália (ICSM, 2009), a UNE 148002:2016 – *Control de la Calidad Posicional en Conjuntos de Datos Espaciales* da Espanha (AENOR, 2016) e a STANAG 2215, criada pela Agência de Padronização da OTAN (Organização do Tratado do Atlântico Norte) (NSA, 2010).

Entretanto, o tamanho amostral para a avaliação da acurácia de produtos cartográficos utilizando feições lineares é um tema pouco pesquisado. Não se tem conhecimento de uma norma que indique o tamanho da amostra. Alguns autores como Mozas-Calvache (2007), Mozas-Calvache e Ariza-López (2010), Ariza-López e Mozas-Calvache (2012), Santos (2015), La Vega, Ariza-López e Mozas-Calvache (2016), Santos et al. (2016), Moreira et al. (2018), Fonseca (2018), Ariza-López, Ruiz-Lendínez e Ureña-Cámara (2018) afirmam que o tamanho amostral na avaliação da acurácia posicional realizada com linhas ainda não foi definido.

O trabalho de Ariza-López et al. (2011) é o único que se tem conhecimento que aborda a questão do tamanho amostral em feições lineares. Estes autores utilizaram a Simulação Monte Carlo para gerar amostras de diferentes comprimentos, para discutir a influência do tamanho da amostra na variabilidade dos resultados de alguns métodos de feições lineares. Entretanto, os autores não levam em conta questões importantes, como a distribuição espacial da amostra, assumindo que esta é sempre adequada e não afeta a validade dos resultados, e o tamanho da área avaliada. Além disso, por não ser o objetivo desse estudo, não há uma indicação do tamanho da amostra a ser utilizado no controle posicional.

Diante do exposto, um estudo que se proponha a estabelecer uma metodologia consistente para determinar o tamanho da amostra de checagem é relevante e necessário para que se possa avançar no tema do uso de feições lineares para a avaliação da acurácia posicional.

Portanto, o objetivo desta pesquisa foi determinar o tamanho amostral das feições lineares utilizadas na avaliação da acurácia posicional, levando em consideração o padrão de distribuição espacial dos elementos de controle e o tamanho da área avaliada.

## 2 MÉTODOS PARA AVALIAÇÃO DA ACURÁCIA POSICIONAL BASEADOS EM FEIÇÕES LINEARES

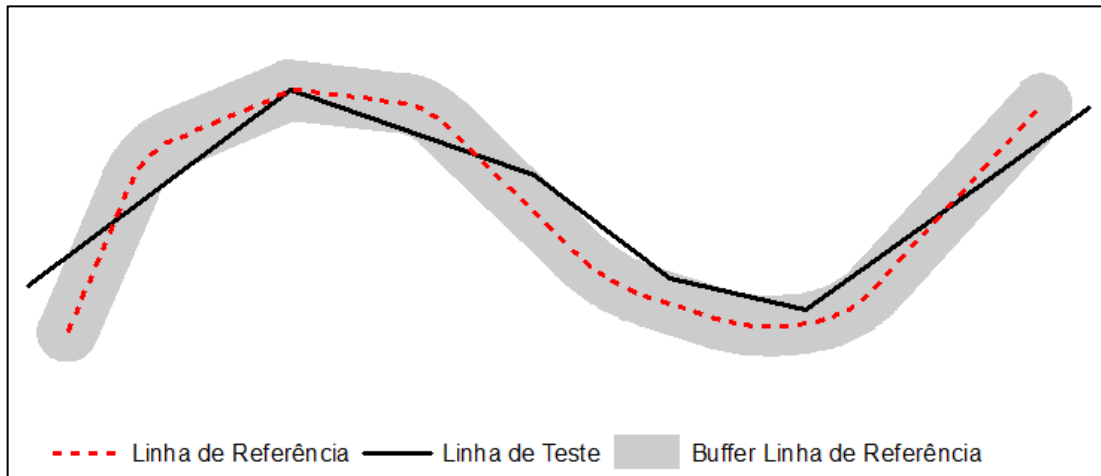
A maioria dos métodos que utilizam elementos lineares para avaliar a acurácia posicional de produtos cartográficos é baseado na comparação de linhas homólogas. Segundo Santos et al. (2015), grande parte destes métodos foi desenvolvido para avaliar os resultados da generalização cartográfica e podem ser divididos em três grandes grupos:

- Grupo de faixa de incerteza que utilizam uma abordagem determinística, como os métodos da Banda Épsilon (também conhecido como Método das Áreas), do *Buffer* Simples e de *Buffer* Duplo;
- Grupo de faixa de incerteza utilizando uma abordagem estocástica, tendo como exemplos a Banda Genérica e o Modelo de Erros Estatístico por Simulação;
- Grupo de métodos determinísticos que realizam análises geométricas, como os métodos da Distância de Hausdorff e da Influência do Vértice.

Este tópico descreve a metodologia de aplicação dos métodos *Buffer* Simples, *Buffer* Duplo, Distância de Hausdorff e Influência do Vértice para avaliação da acurácia posicional planimétrica a partir de feições lineares.

### 2.1 *Buffer* Simples

O método do *Buffer* Simples, conhecido também como *Simple Buffer Overlay Method*, é apresentado no trabalho de Goodchild e Hunter (1997). Este método consiste na geração de um *buffer* a partir da linha de referência, proveniente de uma fonte de melhor acurácia, e determinação da porcentagem do comprimento da linha de teste inclusa dentro do *buffer*, como ilustra a Figura 1 (GOODCHILD e HUNTER, 1997; MOZAS-CALVACHE e ARIZA-LÓPEZ, 2011; SANTOS et al., 2015). Esta porcentagem de inclusão é utilizada para se obter uma medida de discrepância posicional dentre as duas linhas.

Figura 1 - Método do *Buffer* Simples.

Fonte: Adaptado de Santos et al. (2015).

A execução do método do *Buffer* Simples é realizada seguindo os seguintes passos (MOZAS-CALVACHE, 2007; SANTOS et al., 2015; ZANETTI et al., 2016):

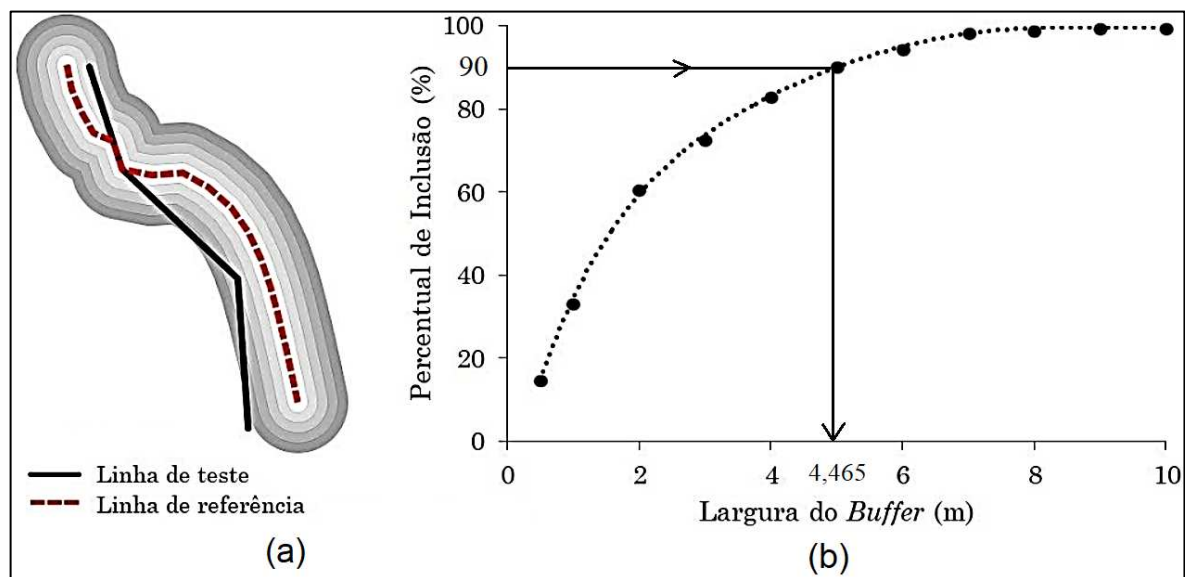
1. Selecionar uma amostra de linhas de teste, no produto em análise;
2. Selecionar linhas homólogas às linhas de teste em um conjunto de referência, com maior acurácia;
3. Calcular o comprimento de cada linha *i* da amostra de teste ( $CL_T$ );
4. Gerar um *buffer* (faixa de incerteza) de largura  $X$  a partir das linhas de referência;
5. Realizar a interseção entre as linhas de teste e os *buffers* das linhas de referência;
6. Calcular o comprimento para cada linha de teste *i* resultante dessa interseção ( $CL_x$ ); e
7. Calcular a porcentagem de inclusão de cada linha de teste dentro da área do *buffer*. Essa porcentagem, dada pela Equação 1, é obtida através da razão entre o comprimento da linha que está incluído no *buffer* e o comprimento total da linha de teste.

$$p(X)_i = \frac{CL_{x_i}}{CL_{T_i}} \quad (1)$$

Como se pode perceber, esta equação fornece uma porcentagem, não uma medida métrica. Diante disso, Goodchild e Hunter (1997) traçam um paralelo com as

medidas utilizadas na avaliação da acurácia posicional por meio de pontos. Segundo estes autores, espera-se que a maioria das aplicações esteja interessada em um valor de  $X$  para uma determinada porcentagem de inclusão  $p(X)$ . Em outras palavras, o interesse está na largura do *buffer* que gera  $p(X)$ . Ainda segundo estes autores, isto pode ser realizado por meio de um processo iterativo, podendo ser os percentis 90 ou 95, que geralmente são utilizados como base para as normas de avaliação da acurácia posicional de mapas. Este processo iterativo pode ser realizado aumentando-se gradativamente a largura do *buffer* (Figura 2a), a fim de se obter uma distribuição de probabilidade de inclusão da linha de teste no *buffer* da linha de referência (Figura 2b), como representado na Figura 2.

Figura 2 – Porcentagem de inclusão da linha de teste em relação à largura do *buffer*. (a) Aumento da largura do *buffer* e (b) distribuição de probabilidade.



Fonte: Adaptada de Ariza-López e Mozas-Calvache (2005).

A partir da relação entre a porcentagem de inclusão e a largura do *buffer* é possível determinar a largura do *buffer* gerado a partir da linha de referência que garanta a sobreposição mínima de 90% ou 95% da linha de teste. A proposta do método é utilizar esta largura de *buffer* como medida de discrepância posicional. No exemplo da Figura 2a, considerando uma porcentagem de inclusão de 90%, a discrepância fornecida pelo método *Buffer Simple* seria de 4,465 m. Nesta pesquisa, foi utilizado o percentil de 90 para determinação da largura do *buffer*. Este valor

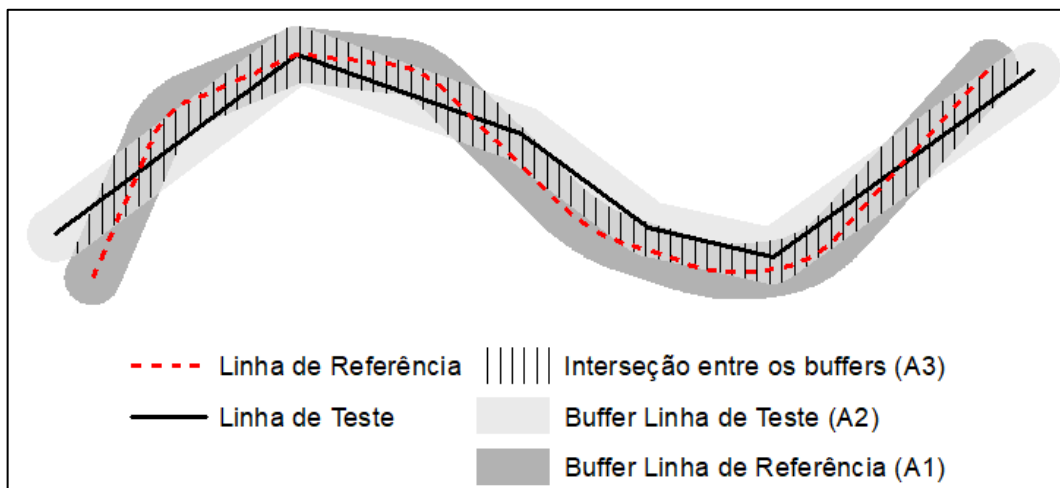
também é utilizado por Cunha et al. (2019), que apresentam uma adaptação do método *Buffer Simples* para a sua utilização seguindo as normas brasileiras.

## 2.2 Buffer Duplo

O método do *Buffer Duplo*, denominado de *Buffer Overlay Statistics* pelos autores, foi proposto por Tveite e Langaas (1999) como um avanço do método do *Buffer Simples*.

Este método consiste em aplicar o modelo da banda de incerteza tanto na linha de referência quanto na linha de teste. Em seguida, deve-se realizar a interseção entre os polígonos resultantes dos *buffers*, como apresentado na Figura 3.

Figura 3 – Método do *Buffer Duplo*.



Fonte: Adaptado de Santos et al. (2015)

Essa discrepância (**dm**), calculada para cada linha de teste **i**, pode ser obtida por meio da Equação 2.

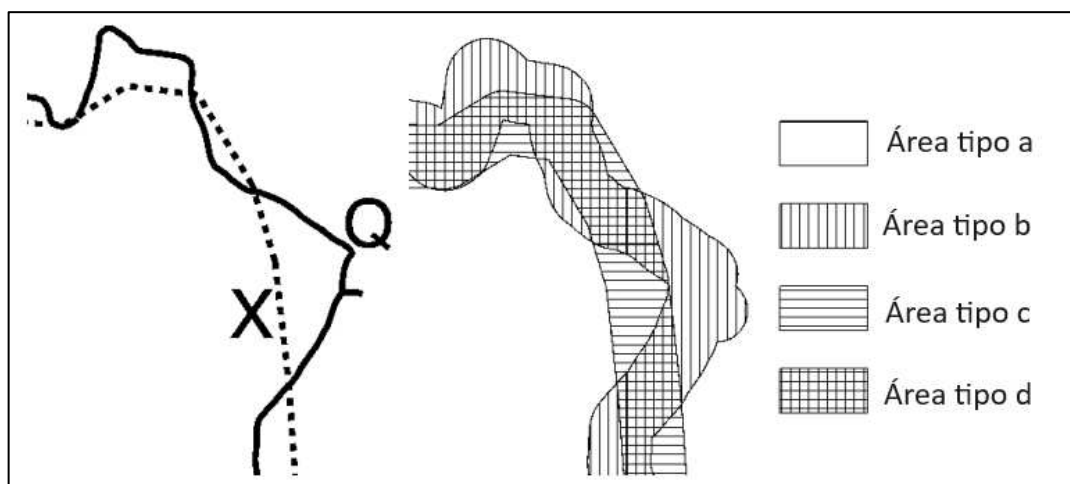
$$dm_i = \pi. x. \frac{[\sum(A_1 - A_3)]_i}{A_{2i}} \quad (2)$$

Nesta equação, ( $A_1 - A_3$ ) representa a área do *buffer* da linha de referência que não faz interseção com o *buffer* da linha de teste. Já  $A_2$  representa a área formada pelo *buffer* da linha teste.

Assim como no caso do método do *Buffer Simples*,  $x$  da Equação 2 representa a largura do *buffer*. Tveite e Langaas (1999) afirmam que a determinação da largura ideal de  $x$  é um processo iterativo. Devem ser aplicados sucessivos *buffers* de largura crescente nas linhas de teste e de referência até que os resultados pareçam estar estabilizados. Para cada largura de *buffer* utilizado nesta etapa deve ser realizada a normalização das áreas. Na aplicação do método do *Buffer Duplo*, como pode ser observado na Figura 4, quatro tipos de área podem surgir (TVEITE e LANGAAS, 1999):

- Áreas tipo **a**: área não contida nos polígonos do *buffer* da linha de referência ( $A_1$ , na Figura 3) nem do *buffer* da linha de teste ( $A_2$ );
- Áreas tipo **b**: área contida no *buffer* da linha de referência, mas não contida no *buffer* da linha de teste;
- Áreas tipo **c**: área contida no *buffer* da linha de teste, mas não contida no *buffer* da linha de referência; e
- Áreas tipo **d**: área contida na interseção entre os *buffers* das linhas de referência e de teste.

Figura 4 - Áreas formadas pela interseção dos *bufers* das linhas de referência (Q) e de teste (X) no método do *Buffer Duplo*.



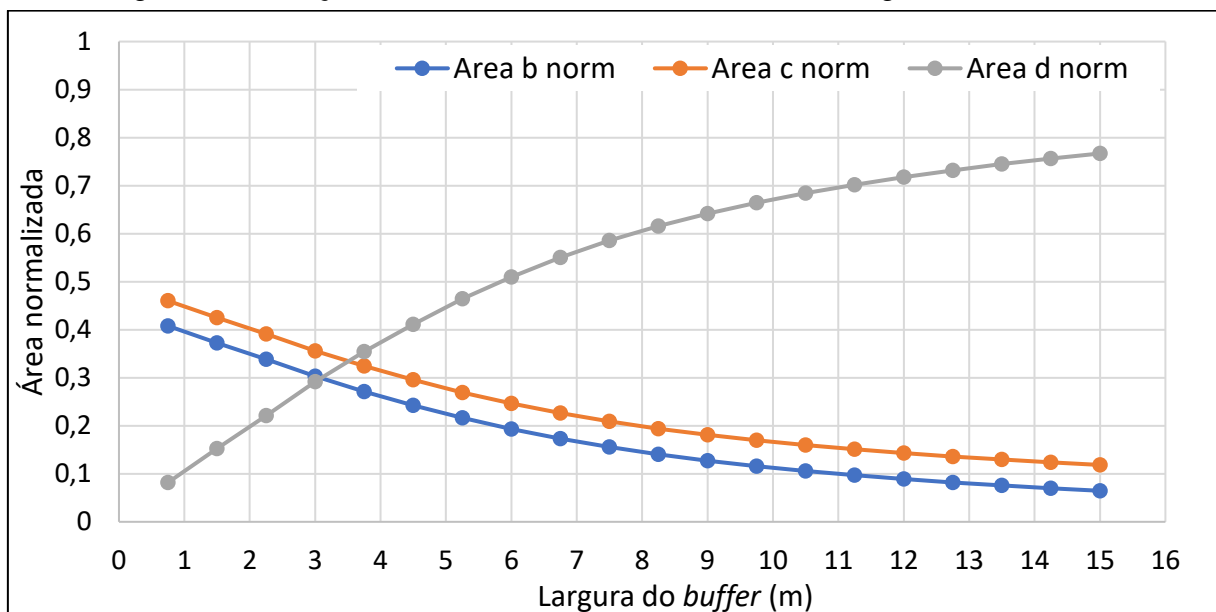
Fonte: Adaptado de Tveite e Langaas (1999)

A normalização destes tipos de área é realizada pela divisão da área em questão pela soma das áreas dos tipos **b**, **c** e **d**. A Equação 3 apresenta o exemplo da normalização da área tipo **b**.

$$\text{Área } b_{\text{normalizada}} = \frac{\text{Área } b}{\text{Área } b + \text{Área } c + \text{Área } d} \quad (3)$$

Após a normalização, pode-se gerar o gráfico da relação entre as larguras dos *buffers* e a inclusão, para cada tipo de área. Um exemplo deste gráfico é apresentado na Figura 5.

Figura 5 - Relação entre as áreas normalizadas e a largura dos *buffers*.



Este gráfico sempre terá dois pontos de interseção, e largura ideal do *buffer* estará contido na região delimitada por estes pontos (TVEITE e LANGAAS, 1999). Neste estudo, para se obter o caso mais restritivo, foi escolhida a maior largura para o *buffer*.

Definida a largura ideal do *buffer* (**x**), pode ser aplicada a Equação 2 para se determinar a discrepância posicional planimétrica.

Para a realização do método do *Buffer Duplo*, sugere-se seguir os passos:

1. Selecionar uma amostra de linhas de teste, no produto em análise;
2. Selecionar linhas homólogas às linhas de teste em um conjunto de referência, com maior acurácia;
3. Aplicar sucessivos *buffers* (faixas de incerteza) de crescente nas linhas de teste e de referência até os resultados se estabilizarem;
4. Obter as áreas de tipo **b**, **c** e **d** para cada *buffer* do passo 3;
5. Normalizar as áreas do passo anterior;
6. Determinar a largura de *buffer* ideal **x** em cada linha **i** da amostra de linhas homólogas de teste e de referência;
7. Calcular a área **A<sub>2</sub>**, gerada pela aplicação do *buffer* em cada linha de teste;
8. Realizar a interseção entre os *buffers* das linhas de teste e de referência;
9. Calcular a área **A<sub>1</sub> – A<sub>3</sub>**, correspondente a não interseção entre os *buffers*; e
10. Calcular a discrepância média (**dm**) para cada linha teste **i** (Equação 2).

### 2.3 Distância de Hausdorff

A distância de Hausdorff é amplamente utilizada para recuperação de informações e analisar a similaridade entre objetos vetoriais ou imagens (ARIZA-LÓPEZ e MOZAS-CALVACHE, 2012).

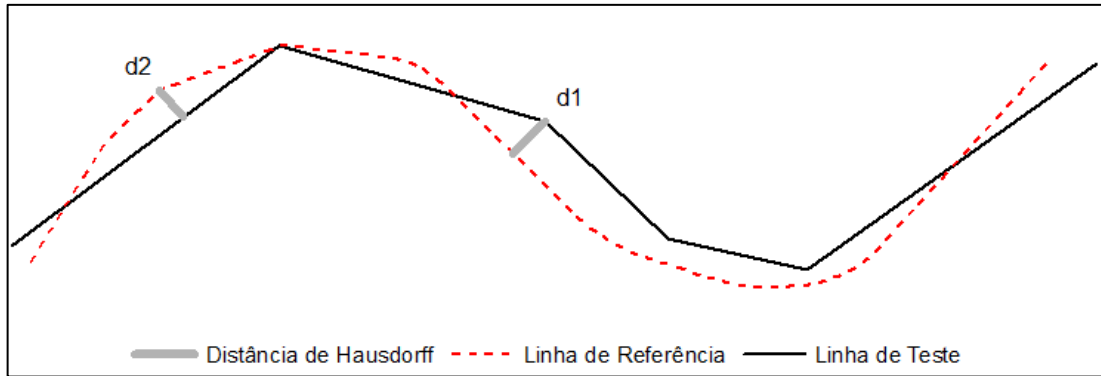
O método, que propõe utilizar a distância de Hausdorff como ferramenta matemática entre um par de linhas para se obter a discrepância posicional, foi desenvolvido por Abbas (1994).

Segundo Santos et al. (2015), para a aplicação desse método, deve-se obter o maior valor entre as menores distâncias de cada linha teste em relação à linha de referência (**d1**), e o maior valor entre as distâncias de cada linha de referência em relação à linha teste (**d2**). A distância de Hausdorff é o maior valor entre **d1** e **d2**.

Entretanto, por utilizar os valores máximos para obter **d1** e **d2** e, conseqüentemente, a distância de Hausdorff, esses autores alertam para o fato de que este método é muito sensível a erros grosseiros ou ruídos. Por isso, surgiram diferentes métodos modificando a distância de Hausdorff para contornar este problema.

Diante disso, Mozas-Calvache (2007) propõe utilizar a média das menores distâncias para se obter **d1** e **d2**, em substituição aos valores máximos. Essa abordagem será utilizada neste trabalho. Um exemplo pode ser visto na Figura 6.

Figura 6 – Método da Distância de Hausdorff.



Fonte: Adaptado de Santos et al. (2015).

Mozas-Calvache (2007) e Santos et al. (2015) apresentam os seguintes passos para a aplicação do método da Distância de Hausdorff:

1. Selecionar uma amostra de linhas de teste, no produto em análise;
2. Selecionar linhas homólogas de uma fonte de melhor acurácia (linhas de referência);
3. Extrair os vértices das linhas de teste e de referência;
4. Calcular a menor distância de cada vértice da linha de teste em relação a todos os segmentos da linha de referência. Então, obter **d1**, que é a média de todas essas menores distância calculadas;
5. Calcular a menor distância de cada vértice da linha de referência em relação a todos os segmentos da linha de teste. Então, obter **d2**, que é a média de todas essas menores distância calculadas; e
6. Calcular a Distância de Hausdorff (**dh**), que é definida como o máximo entre os valores d1 e d2, conforme a Equação 4

$$dh = \max\{d1, d2\} \quad (4)$$

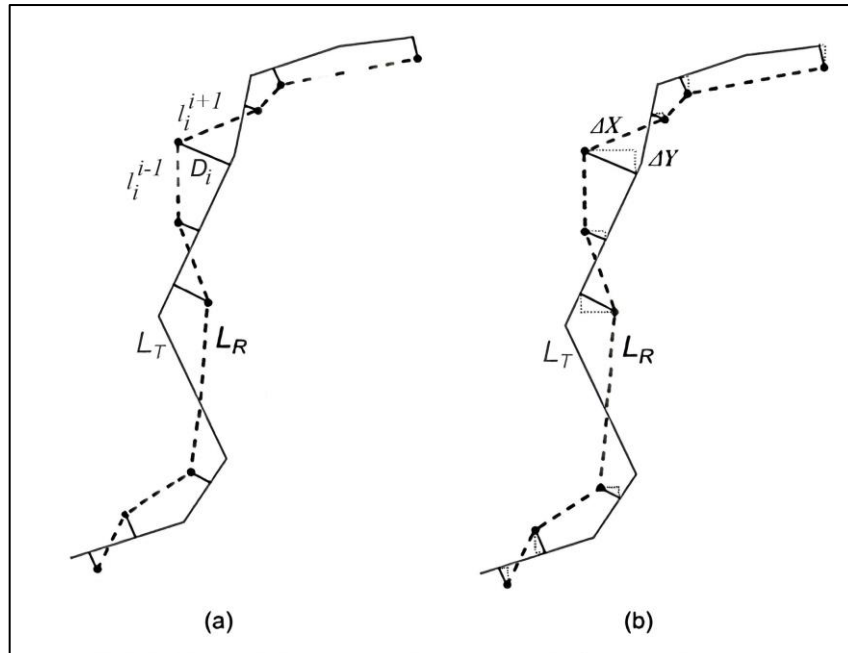
## 2.4 Influência do Vértice

O método da Influência do Vértice, proposto por Mozas-Calvache e Ariza-López (2011), é baseado nos deslocamentos dos vértices e sua influência nos segmentos de linha adjacentes.

Neste método, obtém-se uma medida de discrepância posicional a partir da determinação da distância euclidiana entre os vértices da linha de referência ( $L_R$ ) e a linha de teste ( $L_T$ ). Em seguida, estas distâncias são ponderadas em relação ao comprimento dos segmentos adjacentes ao vértice da linha de referência (SANTOS et al., 2015).

Logo, para cada linha de teste  $i$ , propõe-se calcular a média ponderada usando o comprimento das linhas (peso), para obter o valor da discrepância posicional média ( $D$ ) entre as duas linhas (teste e referência), conforme Equação 5 e Figura 7a.

Figura 7 – Método da Influência do Vértice. (a) Comprimento dos segmentos adjacentes de um vértice; e (b) Componentes  $\Delta X$  e  $\Delta Y$  da distância euclidiana.



Fonte: Mozas e Ariza (2011).

Sendo,

$$D_i = \frac{\left[ \sum_{k=1}^m (ah_k \cdot (l_k^{k-1} + l_k^{k+1})) \right]_i}{2 \cdot CL_{R_i}} \quad (5)$$

Onde,

- $m$ : número de vértices da linha de referência;
- $dh_k$ : distância horizontal mínima entre o vértice  $k$ , de  $L_R$  para  $L_T$ ;
- $l^{k-1}$  e  $l^{k+1}$ : comprimento dos segmentos adjacentes ao vértice  $k$ , em  $L_R$ ; e
- $CL_R$ : comprimento da linha de referência.

De acordo com Santos et al. (2015), para aplicação do método da Influência do Vértice deve-se seguir os seguintes passos:

1. Selecionar uma amostra de linhas de teste, no produto em análise;
2. Selecionar linhas homólogas de uma fonte de melhor acurácia (linhas de referência);
3. Calcular o comprimento da linha de referência ( $CL_R$ ) e extrair os vértices de cada linha  $i$  de referência;
4. Calcular a menor distância ( $dh_k$ ) para cada vértice  $k$  da linha referência ( $L_R$ ) em relação a qualquer segmento da linha teste ( $L_T$ );
5. Calcular os comprimentos dos segmentos de retas adjacentes anteriores ( $l^{k-1}$ ) e posteriores ( $l^{k+1}$ ) aos vértices  $k$  da linha  $i$  de referência; e
6. Calcular a discrepância média ( $D_i$ ), segundo a Equação 5, para cada linha de referência  $i$ .

### 3 SIMULAÇÃO MONTE CARLO

A simulação pode ser definida como a construção de um modelo matemático capaz de reproduzir as características de um fenômeno, sistema ou processo, a fim de obter informações ou resolver problemas (RÍOS, RÍOS e MARTÍN, 1997; ROSS, 2006). O seu principal objetivo é ajudar a entender a relação entre valores reais e estimados, a fim de obter conhecimento empírico sobre determinado parâmetro. (ARIZA-LÓPEZ, RUIZ-LENDÍNEZ e UREÑA-CÁMARA, 2018). A Simulação Monte Carlo é um dos métodos mais utilizados para realizar simulações.

O Método Monte Carlo, ou Simulação Monte Carlo, foi inicialmente utilizado para solucionar problemas de difusão de nêutrons no Projeto Manhattan, para o desenvolvimento da bomba atômica em 1944, na Segunda Guerra Mundial (KALOS e WHITLOCK, 2009; ROFATTO, MATSUOKA e KLEIN, 2017).

Seu nome, dado pelos cientistas John Von Neumann e Stanislaw Ulam, é baseado no Cassino de Monte Carlo, no principado de Mônaco, e é inspirado no aspecto aleatório e repetitivo das roletas, semelhante às características do método (SILVA, 2014).

Esse método é utilizado para solucionar problemas que envolvem um grande número de variáveis, vários tipos de distribuições estatísticas e modelos não-lineares complexos (ROBERT e CASELLA, 2013). Por isso ele apresenta diversas aplicações, nas mais diversas áreas. O Método Monte Carlo pode ser utilizado em áreas como: biologia, ciência política, estatística aplicada, física médica, física teórica, física quântica, entre outras (ROFATTO, MATSUOKA e KLEIN, 2017). Nas Informações Espaciais, exemplos de aplicações da Simulação Monte Carlo podem ser encontradas para simular voos fotogramétricos e o imageamento por satélites (SILVA e DALMOLIN, 2002), para a modelagem espacial (ORTIZ, FELGUEIRAS e RENNÓ, 2007), na busca de um método mais eficiente que os polígonos de Thiessen para a estimativa da chuva diária medida por satélite (GUELFÍ e LÓPEZ-VÁZQUEZ, 2015), na avaliação da precisão da topologia de redes triangulares irregulares (PENHA e COELHO, 2015), na análise da confiabilidade de redes geodésicas (ROFATTO, MATSUOKA e KLEIN, 2017), na análise das incertezas envolvidas na classificação da cobertura do solo (REIS et al., 2017), entre outros. No Controle de Qualidade Cartográfica, a simulação foi abordada em alguns trabalhos em termos de feições pontuais, tais como em Ariza-López (2002a), Ariza-López (2002b), Nero (2005) e Nero et al. (2017), por exemplo.

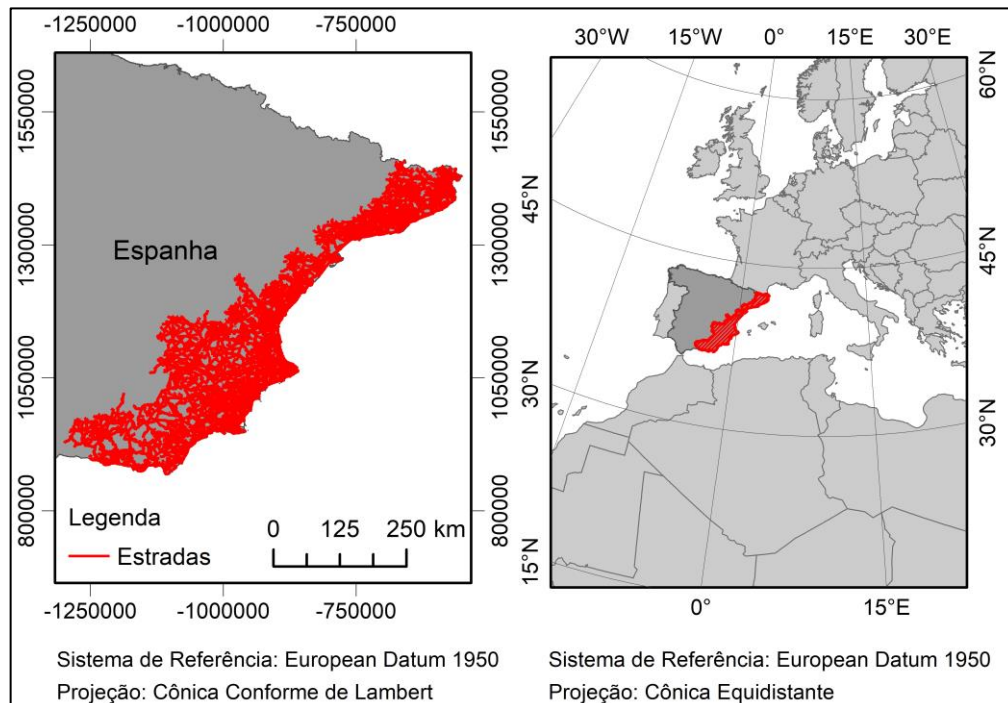
A ideia geral da Simulação Monte Carlo é gerar uma amostragem aleatória a partir da população original e conduzir um grande número de experimentos. Em cada um destes experimentos, os possíveis valores das variáveis aleatórias de entrada são amostrados. Em seguida, os valores da variável de saída são calculados por meio de uma função, que depende da aplicação dada ao método. O resultado do método é obtido por meio de parâmetros estatísticos (como média, desvio-padrão, variância, erro padrão, etc.) e pode ser utilizado para a análise estatística (SILVA, 2014; ROFATTO, MATSUOKA e KLEIN, 2017). Geralmente, nesse processo de simulação, para cada experimento, são sugeridas pelo menos 1000 iterações (BUSLENKO, 1966; HARDIN e SHUMWAY, 1997; MOZAS-CALVACHE, 2007).

## 4 MATERIAIS E MÉTODOS

### 4.1 Área de Estudo

A área de estudo desta pesquisa, representada na Figura 8, compreende 10 províncias do leste da Espanha: Albacete, Alicante, Almería, Barcelona, Castellón, Girona, Granada, Murcia, Tarragona e Valencia. Com uma área de, aproximadamente, 93268 km<sup>2</sup>, esta região é delimitada pelas latitudes 36° 43' 12" N e 42° 29' 43" N, e longitudes 03° 19' 18" E e 04° 19' 40" W.

Figura 8 - Área de Estudo.



### 4.2 Base de Dados

A base de dados utilizada neste estudo é composta por 46460 km de estradas, obtidas no Instituto Geográfico Nacional (IGN) da Espanha. As feições lineares de referência são provenientes do Mapa Topográfico Nacional na escala de 1:25.000 - MTN25 (INSTITUTO GEOGRÁFICO NACIONAL, 2021a). Já as feições lineares de teste foram obtidas na Base Topográfica Nacional na escala de 1:100.000 - BTN100 (INSTITUTO GEOGRÁFICO NACIONAL, 2021b). Estes dados estão disponíveis em: <<http://centrodedescargas.cnig.es/CentroDescargas/buscadorCatalogo.do>>. Foram

utilizados dados da Espanha porque o IGN possui bases de dados de diversas escalas (até 1/1.000, por exemplo), cobrindo uma grande área.

Ariza-López (2002a), Nero (2005) e DSG (2016) afirmam que o produto de referência deve ser pelo menos três vezes mais acurado que o produto que será avaliado. Portanto, os dados utilizados neste trabalho atendem ao que é recomendado.

### 4.3 Metodologia

Após a obtenção dos dados, foi realizada a sua preparação antes do processo de simulação. Essa etapa foi necessária para garantir que os dados de referência e de teste estivessem consistentes topologicamente, e que as análises ocorressem entre feições homólogas.

Inicialmente foi feita uma análise espacial para investigar quais feições de teste não possuíam uma feição de referência próxima, o que poderia indicar que não haveria um homólogo. Esta análise de proximidade foi realizada por meio de *buffers*, considerando uma distância de 100 m. Este valor provém do PEC-PCD da classe D da ET-CQDG para a escala de 1:100.000 (DSG, 2016). Todas as feições de teste que não estavam contidas no *buffer* das feições de referência foram excluídas. Em seguida, foi realizado um procedimento semelhante, mas avaliando a proximidade das linhas de referência em relação às linhas de teste.

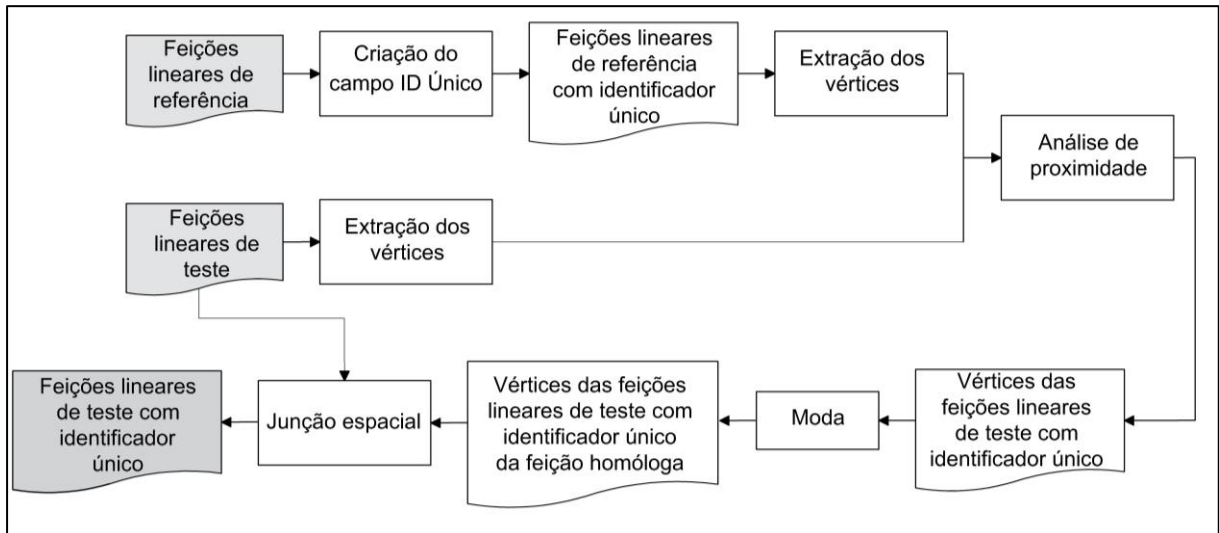
Esta análise espacial não contempla todas as possibilidades para identificação de feições não homólogas. As feições dentro do *buffer*, mesmo sem um par homólogo, seriam preservadas. Dessa forma, foi necessário realizar a edição manual para excluir estas linhas.

O passo seguinte foi aplicar topologia. As regras topológicas utilizadas visaram garantir que as feições fossem divididas nos cruzamentos e não permitir a presença de pseudonós, para assegurar a conectividade das linhas fora dos cruzamentos.

Por fim, a última etapa de preparação dos dados envolveu atribuir a cada feição linear da área de estudo (Figura 8) um identificador único. O objetivo deste identificador é garantir que a avaliação da acurácia posicional seja realizada entre os pares homólogos das feições lineares. Caso a discrepância posicional planimétrica

não seja obtida entre linhas homólogas, as análises poderiam estar equivocadas e não retratariam a realidade dos dados. A metodologia utilizada para esta etapa está representada no fluxograma da Figura 9.

Figura 9 - Metodologia para obtenção do identificador único.



Para que as linhas homólogas possuíssem o mesmo identificador (denominado 'ID Único', neste estudo), a ideia central foi criar um campo de identificação único e sequencial nas feições de referência e transferir, por meio de relacionamentos espaciais, este identificador para as feições de teste. Inicialmente foram extraídos os vértices de cada linha de referência e de teste. Em seguida, foi necessário identificar, para cada vértice de teste, qual vértice da linha de referência estava mais próximo. Este processo transferiu o 'ID Único' dos vértices das feições de referência para os vértices das feições de teste.

Devido a alguns fatores, como a grande densidade de estradas em algumas regiões, o formato das linhas que as representam, e a própria discrepância posicional inerente a cada feição linear da base de dados, nem todos os vértices de uma linha de teste apontavam a mesma linha de referência como sendo a mais próxima e, portanto, sua homóloga. Diante disto, foi necessário adotar um critério para decidir qual seria a linha homóloga. Para isto, foi utilizada a medida estatística da moda. Ou seja, foram consideradas feições homólogas, com o mesmo identificador único, aquelas cuja maioria dos vértices estivesse mais próxima. Para aplicar esta sequência de procedimentos, foi elaborado um código no *software* R Studio (R CORE TEAM, 2023).

Por fim, para que a identificação dos vértices das feições de teste fosse transferida para suas linhas correspondentes, foi realizada uma junção espacial. A validação foi realizada por meio de outra análise de proximidade e por inspeção visual. Esta análise de proximidade, realizada entre os vértices médios das linhas de teste e de referência, buscou indicar o 'ID Único' dos vértices médios das linhas de teste mais próximo a cada linha de referência, e vice-versa. Uma não correspondência entre os 'ID Único' de referência e de teste indicaria um possível erro na atribuição dos identificadores. Além disso, foi realizada uma análise visual em busca de possíveis inconsistências, como identificadores duplicados ou ausentes. Utilizando esta sequência metodológica, mais de 98% das feições presentes na base de dados tiveram seu identificador único atribuído de forma correta.

Uma etapa intermediária, realizada após a preparação dos dados e antes do processo de simulação, envolveu a aplicação do algoritmo Douglas-Peucker (DOUGLAS e PEUCKER, 1973) nas duas bases de dados. Neste procedimento foi utilizada uma tolerância de 1 mm, ou seja, foram eliminados os vértices cujo impacto na geometria da linha não fosse superior a 1 mm. Este valor de tolerância foi utilizado porque o objetivo desta etapa foi diminuir o esforço computacional e tempo de processamento, tendo em vista a grande quantidade de experimentos exigidos pela Simulação Monte Carlo, e não alterar posicionalmente as linhas.

Após estas etapas, pôde ser realizada a simulação. O objetivo da simulação foi gerar uma grande quantidade de amostras aleatórias e inferir, baseado na comparação com a população, sobre qual tamanho amostral mínimo seria representativo, sendo o resultado relacionado com o tamanho da área e o padrão de distribuição espacial.

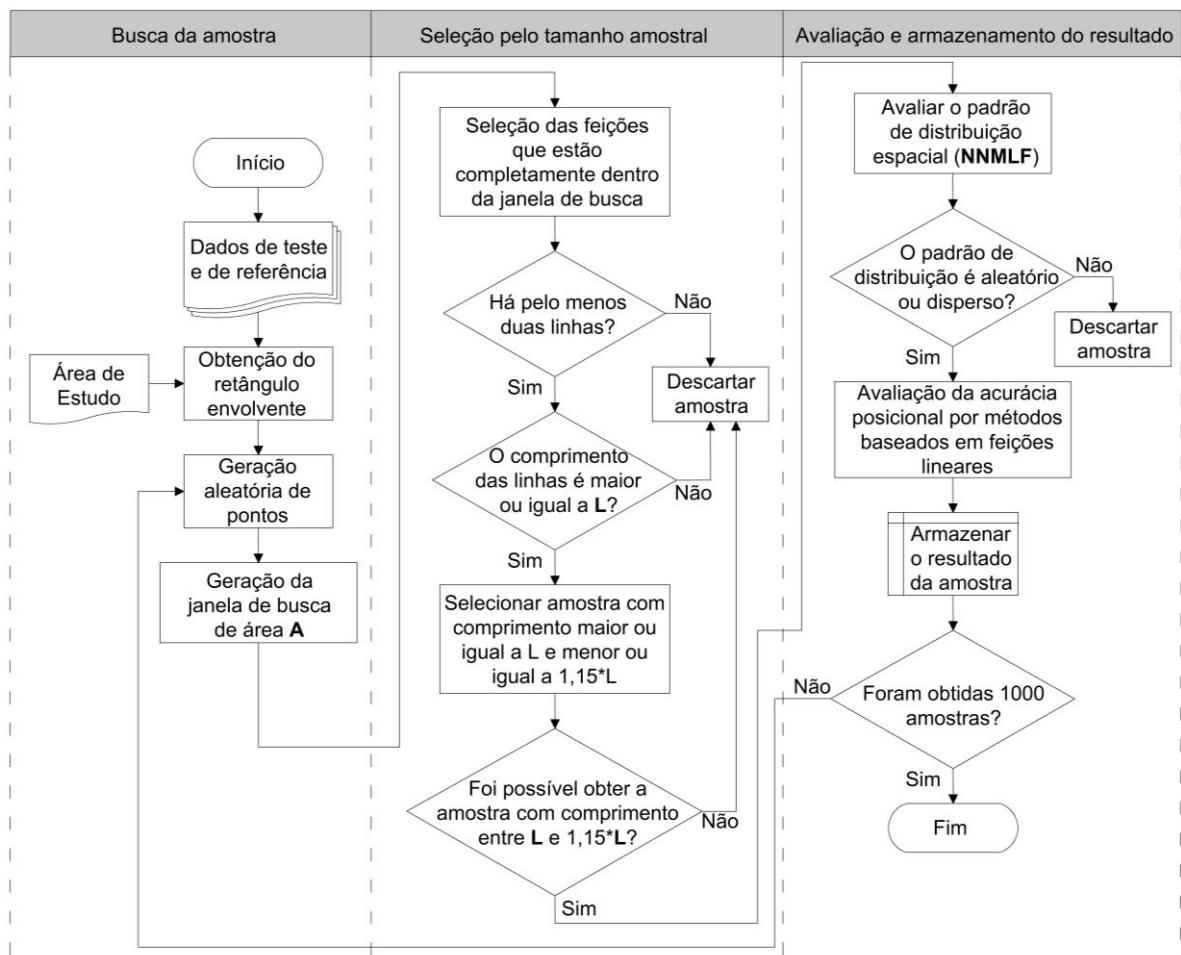
Nesta pesquisa foram utilizados 17 tamanhos de área: 1; 2,5; 5; 7,5; 10; 15; 20; 25; 50; 75; 100; 125; 150; 175; 200; 225 e 250 km<sup>2</sup>. Esta grande diferença entre os tamanhos de área analisados interfere na escolha da população. Não seria razoável comparar as amostras provenientes das áreas de 1 e 250 km<sup>2</sup> com a mesma população, por exemplo. Diante disso, a partir das bases de dados (Figura 8) foram selecionadas feições lineares de referência e de teste para comporem 17 populações, cada uma servindo de base para o processo de simulação em um tamanho de área. A ideia é que houvesse uma proporcionalidade, em termos de comprimento, entre os

tamanhos da amostra e da população. Caso contrário, seria inviável se obter um tamanho de amostra que fosse representativo da população.

A definição da relação entre os comprimentos da população e da amostra ocorreu de maneira empírica. Neste estudo ficou determinado que o comprimento da população corresponderia a 240 vezes o tamanho amostral máximo. Em outras palavras, o tamanho amostral máximo correspondeu a, aproximadamente, 0,40% do comprimento da população. Como comparação, Ariza-López et al. (2011) utilizaram uma população composta por 1254 estradas em seu estudo, cujo comprimento total foi 1210 km. Esta população foi comparada com amostras cujo comprimento máximo foi de 20 km. Neste caso, a amostra representou, no máximo, 1,65% da população, em termos de comprimento.

Definido o critério para obtenção da população, as amostras foram geradas pela Simulação Monte Carlo, cuja sequência metodológica está representada na Figura 10.

Figura 10 - Simulação Monte Carlo.



Em partes, o processo de simulação realizado nesta pesquisa é semelhante ao utilizado no trabalho de Ariza-López et al. (2011), que utilizaram as rodovias da Andaluzia, na costa sul da Espanha. As principais diferenças residem no fato de que nesse estudo foi levado em consideração o padrão de distribuição espacial das feições de controle e o tamanho da área avaliada.

A simulação e os parâmetros utilizados podem ser descritos nos seguintes passos:

1. Selecionar os dados de teste e de referência, que correspondem a população que dará origem às amostras.
2. Obter o retângulo envolvente dos dados.
3. Gerar pontos aleatórios dentro da região do passo anterior. Esta é uma amostragem aleatória simples, para garantir que as feições selecionadas para compor a amostra possuam a mesma probabilidade de serem escolhidas. Além disso, tendo em vista que a área era irregular, nesta etapa, para que a geração dos pontos ocorresse somente onde realmente havia feições, foi inserido um *shapefile* representando a área de estudo.
4. Gerar janelas de busca de área **A**. Estas janelas são polígonos quadrados cujos centros são os pontos do passo anterior. Os tamanhos da área utilizados são aqueles mencionados anteriormente: 1; 2,5; 5; 7,5; 10; 15; 20; 25; 50; 75; 100; 125; 150; 175; 200; 225 e 250 km<sup>2</sup>.

A área da região avaliada é um fator importante quando se trata do tamanho da amostra. Pois, espera-se que quanto maior a área a ser avaliada, maior o tamanho da amostra das feições de controle necessária.

5. Para cada janela de busca, selecionar somente as linhas que se encontram completamente dentro da respectiva janela.
6. Verificar se há pelo menos duas linhas selecionadas no passo 5. Esta etapa é necessária para garantir que se possa aplicar o Nearest Neighbor Method for Linear Features (NNMLF) de primeira ordem, pelo menos. Caso haja somente uma feição linear na janela de busca, esta amostra é descartada.
7. Verificar se o somatório do comprimento das feições presentes na janela de busca é igual ou superior ao tamanho amostral analisado (**L**). Caso isto não ocorresse, esta amostra deveria ser descartada. Para iniciar o processo de simulação, foi definido um tamanho amostral mínimo (**L = L<sub>min</sub>**) que fosse

relacionado com o tamanho da área analisada. Para uma área quadrada **A**, o tamanho mínimo da amostra considerado foi definido como sendo igual ao comprimento de um dos lados desta área, conforme indica a Equação 6.

$$L_{min} = \sqrt{A} \quad (6)$$

8. Selecionar aleatoriamente um conjunto de feições para compor a amostra de tamanho **L**. Tendo em vista a impossibilidade de se obter um tamanho amostral cujo comprimento fosse exatamente igual a **L**, foi admitida uma tolerância de 15% para a última linha incluída na amostra. Desta forma, todas as amostras deveriam possuir comprimento entre **L** e 1,15\***L**. A função desenvolvida para realizar esta tarefa, escrita na linguagem R, se encontra no Apêndice E.
9. Avaliar o padrão de distribuição espacial da amostra. Para esta etapa, foi utilizado o método proposto no tópico anterior (NNMLF). Foram admitidas somente amostras cujas feições apresentaram um padrão de distribuição espacial aleatório ou disperso. As amostras agrupadas são descartadas.
10. Realizar a avaliação da acurácia posicional planimétrica da amostra, seguindo as metodologias descritas no Tópico 2, aplicando os métodos baseados em feições lineares: *Buffer* Simples (BS), *Buffer* Duplo (BD), Distância de Hausdorff (DH) e Influência do Vértice (IV). Estes métodos também foram implementados em linguagem R e se encontram nos Apêndices A, B, C e D, respectivamente. Sendo que os três primeiros métodos desenvolvidos em programação paralela, para melhor desempenho. Este procedimento também requer o cálculo das estatísticas de discrepância para cada método e o armazenamento dos resultados obtidos. O Apêndice G demonstra como foi implementado o *script* em R para o cálculo das estatísticas.
11. Repetir os passos de 3 a 11 até que se obtenham 1000 amostras para o tamanho amostral **L** e tamanho de área **A**. A utilização de 1000 amostras é indicada em trabalhos como Buslenko (1966), Hardin e Shumway (1997) e Mozas-Calvache (2007).

12. Incrementar o tamanho amostral em  $\Delta L$ . Através de experimentos empíricos, se concluiu que um valor razoável para  $\Delta L$  seria a metade do tamanho amostral mínimo, conforme a Equação 7.

$$\Delta L = 0,5 * L_{min} \quad (7)$$

13. Repetir os passos de 3 a 13 para 10 tamanhos amostrais. Caso a décima amostra não seja representativa da população (isto é, a hipótese nula seja rejeitada em algum teste de hipóteses), repetir até que algum tamanho de amostra seja representativo. Entretanto, espera-se que, para a maioria dos casos, 10 tamanhos de amostra sejam suficientes. Portanto, espera-se que o tamanho amostral máximo ( $L_{max}$ ) para determinada área ( $A$ ) seja aquela dada pela Equação 8.

$$L_{max} = L_{min} + 9 * \Delta L \quad (8)$$

Como  $\Delta L$  é dado pela Equação 7, o tamanho amostral máximo ( $L_{max}$ ) também pode ser escrito segundo a Equação 9.

$$L_{max} = L_{min} + 9 * 0,5 * L_{min} \quad (9)$$

Substituindo  $L_{min}$  pela Equação 6, tem-se a Equação 10.

$$L_{max} = 5,5 * \sqrt{A} \quad (10)$$

Este é o tamanho amostral máximo esperado para o processo de simulação de uma área de tamanho  $A$ . Este valor foi aquele que serviu de parâmetro para se obter a população de cada área analisada neste estudo, antes do início da simulação. Dessa forma, para se inferir sobre o tamanho amostral de feições lineares utilizadas no controle posicional presentes em uma área de tamanho  $A$ , este estudo propôs que o tamanho da população, em termos de comprimento, deveria ser 240 vezes o tamanho amostral máximo (Equação 10).

14. Repetir os passos de 1 a 14 para todos os tamanhos de área **A** mencionados anteriormente.

Esta sequência de procedimentos para a implementação da Simulação Monte Carlo se encontra no código apresentado no Apêndice F.

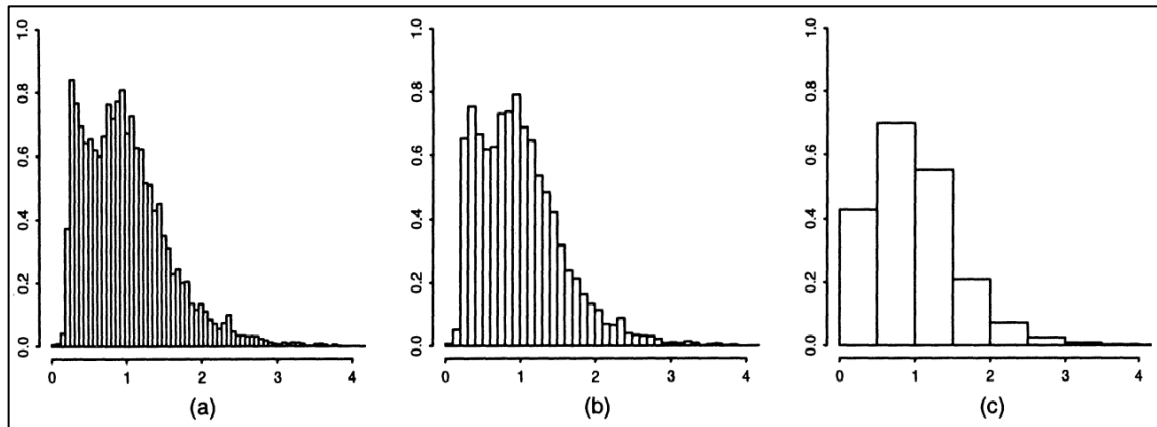
Após a realização do processo de simulação, a última etapa da metodologia envolveu a aplicação do teste Kolmogorov-Smirnov. Este teste pode ser utilizado para avaliar a similaridade entre a Função de Distribuição Acumulada da População (FDP), designada por **F(x)**, e a Função de Distribuição Acumulada Empírica (FDE) ou amostral, denominada **F<sub>e</sub>(x)**. Em outras palavras, o teste Kolmogorov-Smirnov avalia se **F<sub>e</sub>(x)** é um bom estimador para **F(x)**, considerando determinado nível de significância (BUSSAB e MORETTIN, 2010). No caso deste estudo, cada tamanho amostral deu origem a quatro FDE, uma para cada método baseado em feições lineares, e a variável aleatória **x** correspondeu a discrepância posicional planimétrica.

A hipótese nula (**H<sub>0</sub>**) é que **F(x)** e **F<sub>e</sub>(x)** são iguais para todo **x**. Já a hipótese alternativa é que estas funções são diferentes para algum valor de **x**.

$$\begin{cases} H_0: F(x) = F_e(x), \forall x, \\ H_1: F(x) \neq F_e(x), \text{ para algum } x \end{cases}$$

Definida as hipóteses do teste, a primeira etapa é obter as funções de distribuição acumulada. Scott (1979) e Bussab e Morettin (2010) afirmam que o histograma é um estimador da verdadeira função densidade, que dá origem às Funções de Distribuição Acumulada. Portanto, inicialmente foi necessário obter os histogramas da população e das amostras. Wand (1997) afirma que o parâmetro mais importante de um histograma é a largura da caixa, pois ela determina se o gráfico irá apresentar muitos (Figura 11a) ou poucos detalhes (Figura 11c). A Figura 11b apresenta o caso ideal, com uma suavização intermediária.

Figura 11 - Influência da largura da caixa no histograma.



Fonte: Wand (1997).

Existem diversos trabalhos com propostas para a determinação da largura da caixa, como Sturges (1926), Scott (1979), Freedman e Diaconis (1981) e Wand (1997). Entretanto, uma análise preliminar apontou que a utilização de uma regra geral para determinar a largura da caixa poderia não ser adequada a todos os conjuntos de discrepâncias, podendo gerar histogramas muito suavizados para os dados desta pesquisa. De fato, não seria razoável supor que uma única regra seria válida para conjuntos de dados tão diversos, vindos de diferentes métodos, com diferentes amplitudes, desvios-padrão, RMS (*Root Mean Square*), etc. Desta forma, a determinação da largura da caixa foi realizada de forma empírica, para cada FDP. Entretanto, para evitar que ocorresse uma suavização excessiva (Figura 11c), foi utilizada a largura da caixa ( $h_n$ ) proposta por Scott (1979) (Equação 11) como limite superior deste estudo.

$$h_n = 3,49 * s * n^{-\frac{1}{3}} \quad (11)$$

Onde,

- **s** representa o desvio-padrão dos dados; e
- **n** é o número de caixas.

Construídos os histogramas das populações e das amostras, foram obtidas as frequências relativas e, em seguida, as frequências acumuladas. Sendo que estas últimas representam as Funções de Distribuição Acumulada.

A estatística do teste é dada pela máxima diferença absoluta entre as funções  $F(\mathbf{x})$  e  $F_e(\mathbf{x})$ , conforme Equação 12.

$$D = \max_{1 \leq i \leq n} |F(x_i) - F_e(x_i)| \quad (12)$$

A hipótese nula será rejeitada se **D** for maior que o valor tabelado. Para um nível de significância de 5% ( $\alpha = 0,05$ ) e um número de caixas (**n**) superior a 50, como neste estudo, o **D** crítico (**D<sub>c</sub>**) é dado pela Equação 13 (BUSSAB e MORETTIN, 2010).

$$D_c = \frac{1,358}{\sqrt{n}} \quad (13)$$

## 5 RESULTADOS E DISCUSSÕES

Após a obtenção e preparação da base de dados, foram selecionadas as feições lineares de referência e de teste que iriam compor cada uma das 17 populações. Sendo que cada população foi obtida com o objetivo de ser proporcional ao tamanho da amostra máximo analisado em cada tamanho de área. A Figura 12 apresenta estas populações.

Figura 12 – Populações proporcionais aos tamanhos máximos das amostras. (a) 1 km<sup>2</sup>; (b) 2,5 km<sup>2</sup>; (c) 5 km<sup>2</sup>; (d) 7,5 km<sup>2</sup>; (e) 10 km<sup>2</sup>; (f) 15 km<sup>2</sup>; (g) 20 km<sup>2</sup>; (h) 25 km<sup>2</sup>; (i) 50 km<sup>2</sup>; (j) 75 km<sup>2</sup>; (k) 100 km<sup>2</sup>; (l) 125 km<sup>2</sup>; (m) 150 km<sup>2</sup>; (n) 175 km<sup>2</sup>; (o) 200 km<sup>2</sup>; (p) 225 km<sup>2</sup> e (q) 250 km<sup>2</sup>.

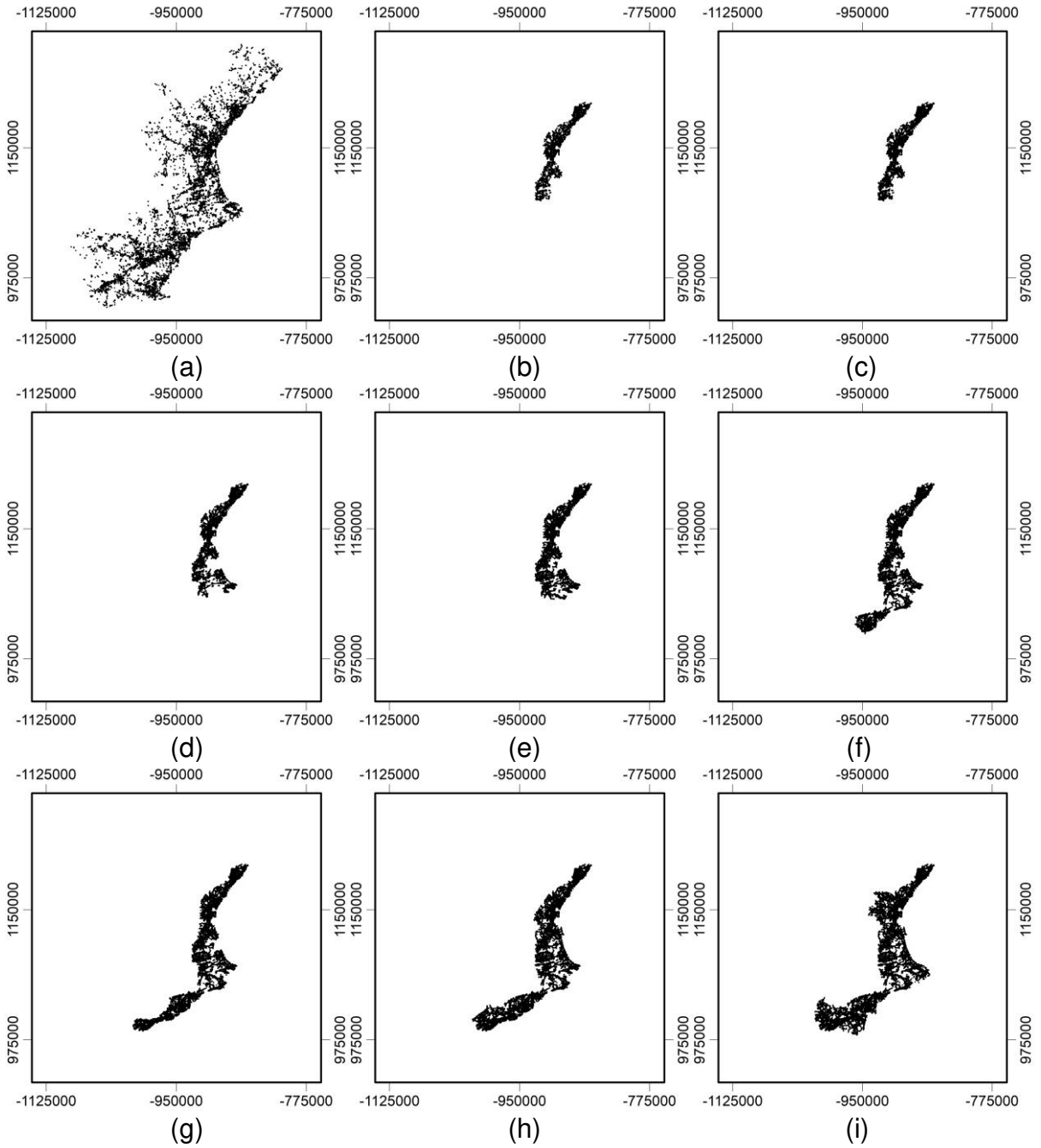
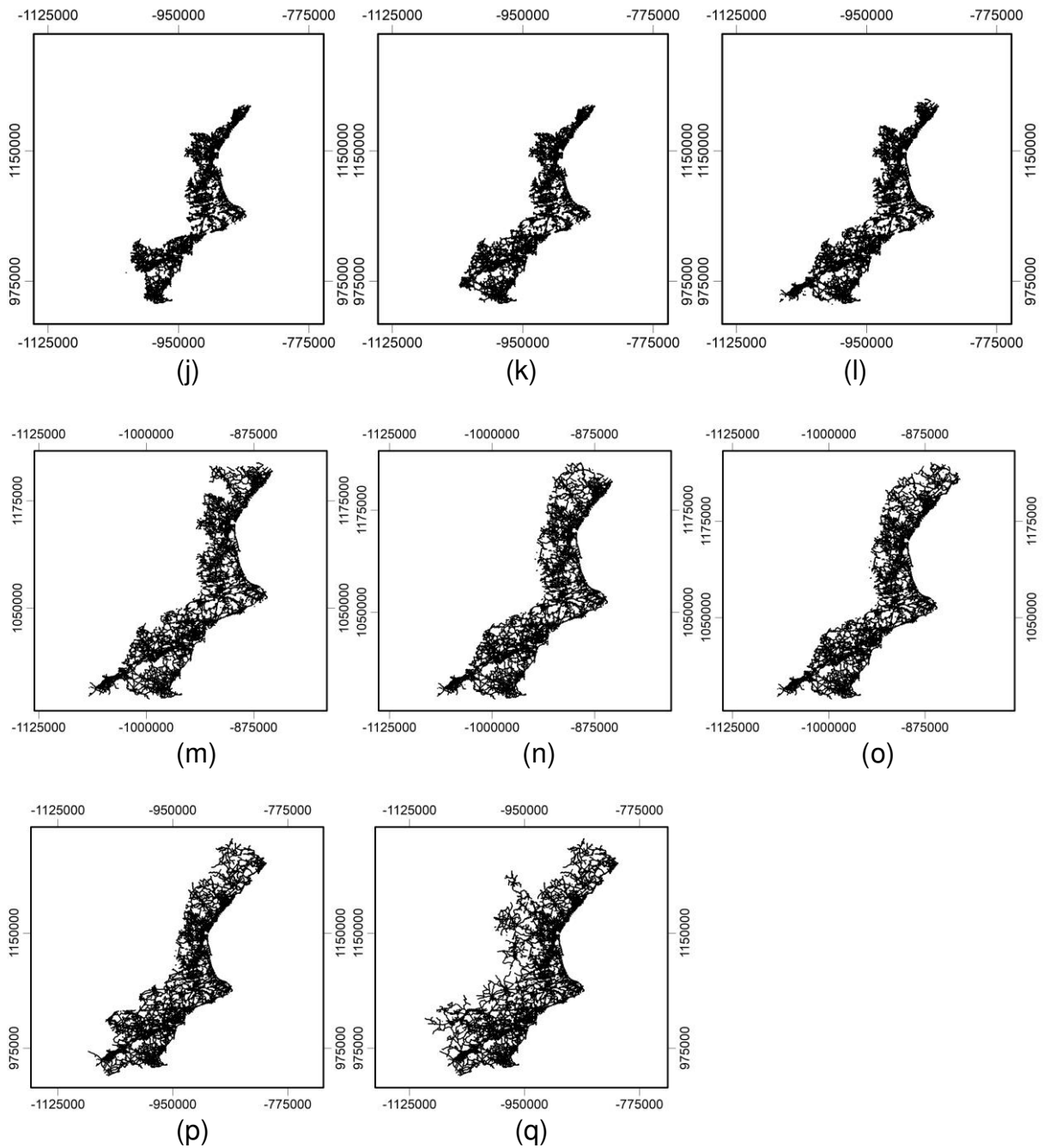


Figura 12 – Populações proporcionais aos tamanhos máximos das amostras. (a) 1 km<sup>2</sup>; (b) 2,5 km<sup>2</sup>; (c) 5 km<sup>2</sup>; (d) 7,5 km<sup>2</sup>; (e) 10 km<sup>2</sup>; (f) 15 km<sup>2</sup>; (g) 20 km<sup>2</sup>; (h) 25 km<sup>2</sup>; (i) 50 km<sup>2</sup>; (j) 75 km<sup>2</sup>; (k) 100 km<sup>2</sup>; (l) 125 km<sup>2</sup>; (m) 150 km<sup>2</sup>; (n) 175 km<sup>2</sup>; (o) 200 km<sup>2</sup>; (p) 225 km<sup>2</sup> e (q) 250 km<sup>2</sup>.



As principais características de cada população são apresentadas na Tabela 1.

Tabela 1 - Características de cada população.

| Figura 12 | Comprimento da população (km) | Comprimento médio das feições (m) | Número de feições | Área (km <sup>2</sup> ) |
|-----------|-------------------------------|-----------------------------------|-------------------|-------------------------|
| a         | 7917,066                      | 578,395                           | 13688             | 33375,117               |
| b         | 2326,939                      | 662,379                           | 3513              | 2444,004                |
| c         | 2814,300                      | 757,551                           | 3715              | 2481,173                |
| d         | 3574,829                      | 767,955                           | 4655              | 3612,391                |
| e         | 4314,130                      | 874,722                           | 4932              | 3911,435                |
| f         | 5554,775                      | 891,331                           | 6232              | 5386,286                |
| g         | 6710,465                      | 887,628                           | 7560              | 6067,671                |
| h         | 7799,085                      | 911,321                           | 8558              | 7302,817                |
| i         | 9247,424                      | 935,406                           | 9886              | 8957,059                |
| j         | 10832,480                     | 946,895                           | 11440             | 10706,098               |
| k         | 12235,244                     | 974,454                           | 12556             | 13221,284               |
| l         | 13382,532                     | 984,154                           | 13598             | 14797,381               |
| m         | 14597,502                     | 1033,598                          | 14123             | 15951,555               |
| n         | 15313,070                     | 1065,999                          | 14365             | 17527,280               |
| o         | 16130,857                     | 1078,771                          | 14953             | 19361,928               |
| p         | 19584,793                     | 1164,929                          | 16812             | 25455,747               |
| q         | 23020,759                     | 1262,241                          | 18238             | 36168,107               |

Analisando a Tabela 1, reitera-se que a decisão de se utilizar uma população para cada simulação foi coerente. Se esta estratégia não fosse utilizada, tamanhos amostrais máximos de 5,5 km (1 km<sup>2</sup>) e 87,0 km (250 km<sup>2</sup>) seriam analisados segundo a mesma população de 46460 km (Figura 8). Dessa forma, provavelmente seria necessário um grande tamanho de amostra para que esta fosse considerada representativa da população. Especialmente para as menores áreas, este não seria um cenário viável de se obter na prática.

Além disso, nesta tabela também é possível observar o grande número de feições que compõe a população destinada a avaliação de amostras dentro da área de 1 km<sup>2</sup>. Isto se deve ao fato de que ficou evidente ao longo deste estudo a falta de feições lineares que, devido ao seu comprimento, poderiam estar contidas nesta área (janela de busca). Devido a isto, buscou-se selecionar apenas as menores linhas disponíveis na área de estudo para compor esta população, maximizando o número de feições que poderiam compor as amostras e se adequassem a área. Esta falta de pequenas feições provavelmente ocorreu devido à escala das linhas de teste (1/100.000). Quanto menor a escala, menos detalhes são representados, sendo que estes detalhes poderiam incluir pequenas linhas.

Sobre estes conjuntos de dados, seguindo as metodologias descritas no Tópico 2, foram aplicados os métodos *Buffer* Simples (BS), *Buffer* Duplo (BD), Distância de Hausdorff (DH) e Influência do Vértice (IV). Os gráficos que representam as discrepâncias posicionais planimétricas obtidas para cada população estão presentes no Apêndice H. Já as tabelas com as estatísticas se encontram no Apêndice I.

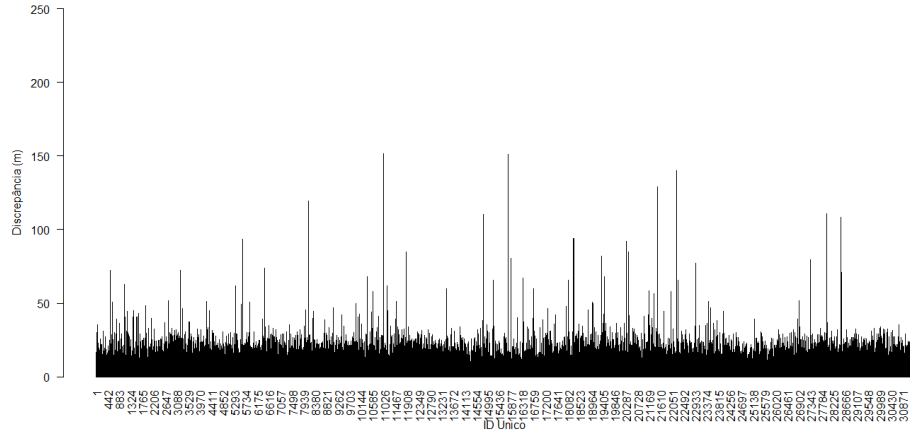
Para ilustrar estas discrepâncias, a Figura 13 traz o resultado da aplicação dos métodos de feição linear em toda base de dados que deu origem às populações (Figura 8). A Tabela 2 apresenta os valores máximos e mínimos, média, desvio-padrão e RMS (*Root Mean Square*) para cada um dos métodos.

Tabela 2 - Estatísticas das discrepâncias da base de dados.

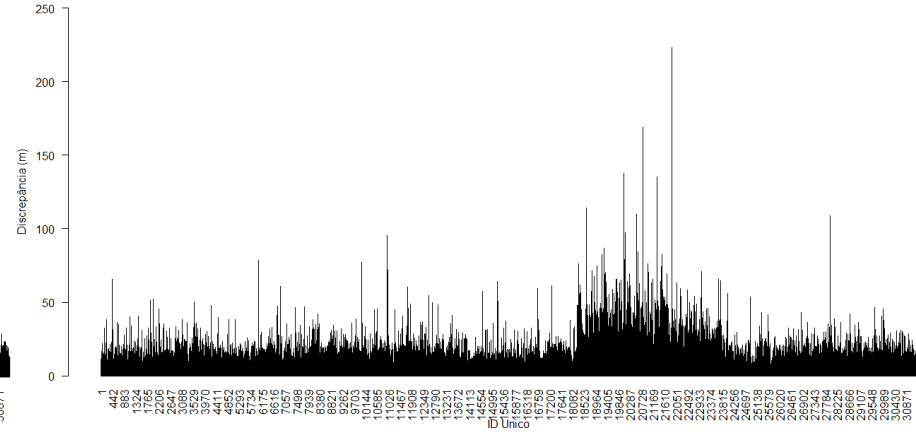
| Estatística (m) | BS      | BD      | DH      | IV     |
|-----------------|---------|---------|---------|--------|
| Mínimo          | 0,142   | 0,000   | 0,044   | 0,092  |
| Máximo          | 151,509 | 223,006 | 167,525 | 61,765 |
| Média           | 10,115  | 8,944   | 7,675   | 5,625  |
| Desvio-Padrão   | 6,284   | 7,577   | 4,957   | 3,437  |
| RMS             | 11,907  | 11,723  | 9,137   | 6,592  |

De modo geral, os métodos *Buffer* Simples e *Buffer* Duplo foram os que apresentaram as maiores discrepâncias. Ariza-López et al. (2011) afirmam que o método *Buffer* Simples superestima as discrepâncias posicionais. Isso porque as distâncias aplicadas para gerar o *buffer* são perpendiculares à linha de referência, o que seria o caso rigoroso. Segundo estes autores, esta faixa de incertezas pode ser mais complexa do que isso. Este mesmo raciocínio pode ser estendido ao *Buffer* Duplo, já que este método também utiliza *buffers*.

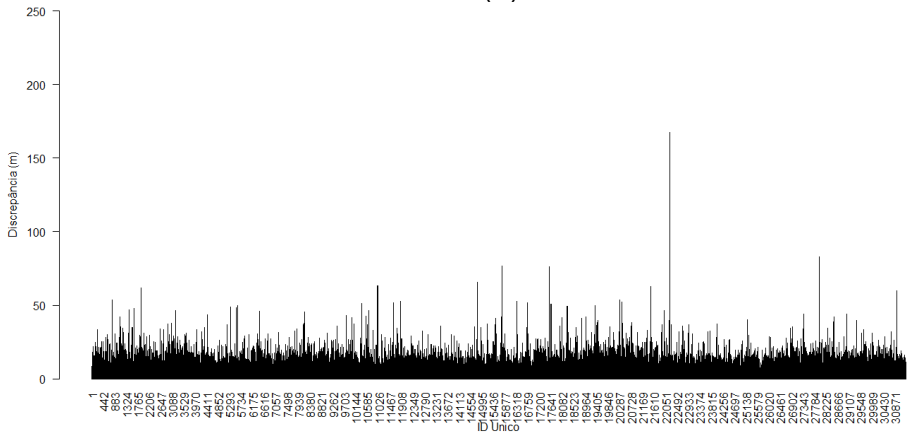
Figura 13 - Discrepâncias posicionais da base de dados. (a) *Buffer Simples*; (b) *Buffer Duplo*; (c) *Distância de Hausdorff* e (d) *Influência do Vértice*.



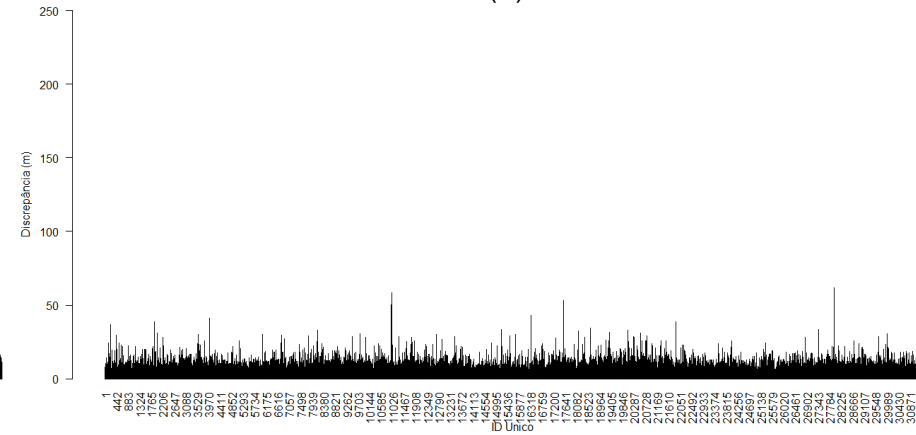
(a)



(b)



(c)

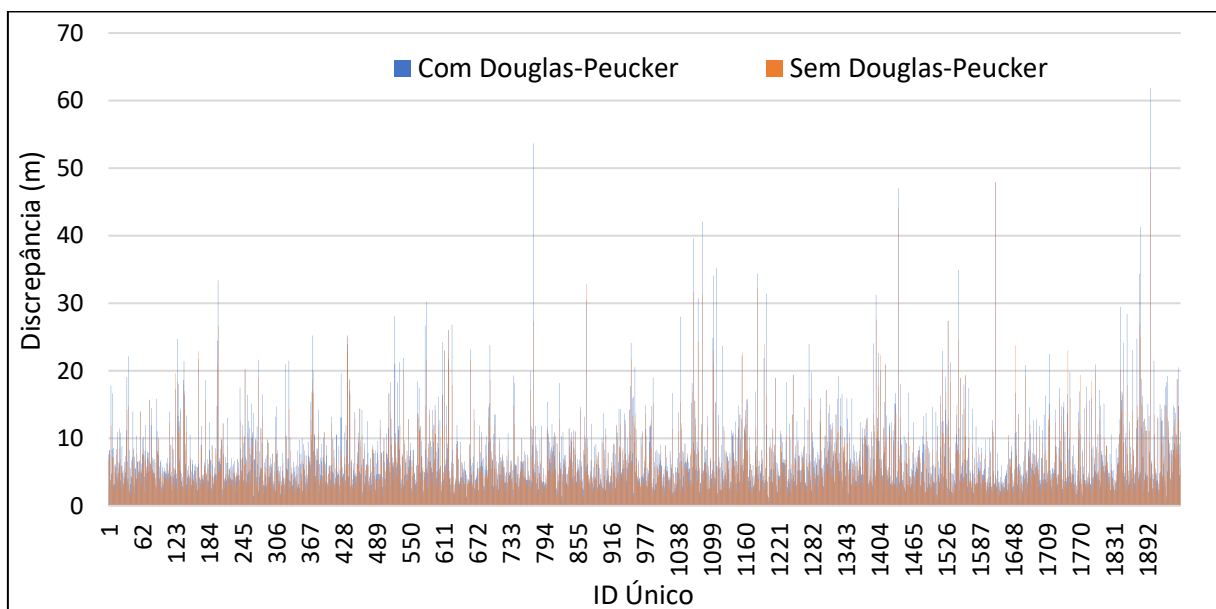


(d)

Como pode ser observado na Figura 13, os métodos baseados na análise da geometria da linha, métodos da Distância de Hausdorff e da Influência do Vértice, foram os que apresentaram as menores discrepâncias. Tendo em vista que foi utilizado o algoritmo Douglas-Peucker nos dados, realizou-se uma análise para verificar se a aplicação deste algoritmo foi o responsável por estes métodos apresentarem discrepâncias inferiores aos métodos que utilizam a faixa de incertezas. Apesar de ter sido utilizada uma tolerância de 1 mm, com o objetivo de não alterar as linhas posicionalmente, este algoritmo elimina vértices desnecessários, o que pode afetar o resultado destes métodos.

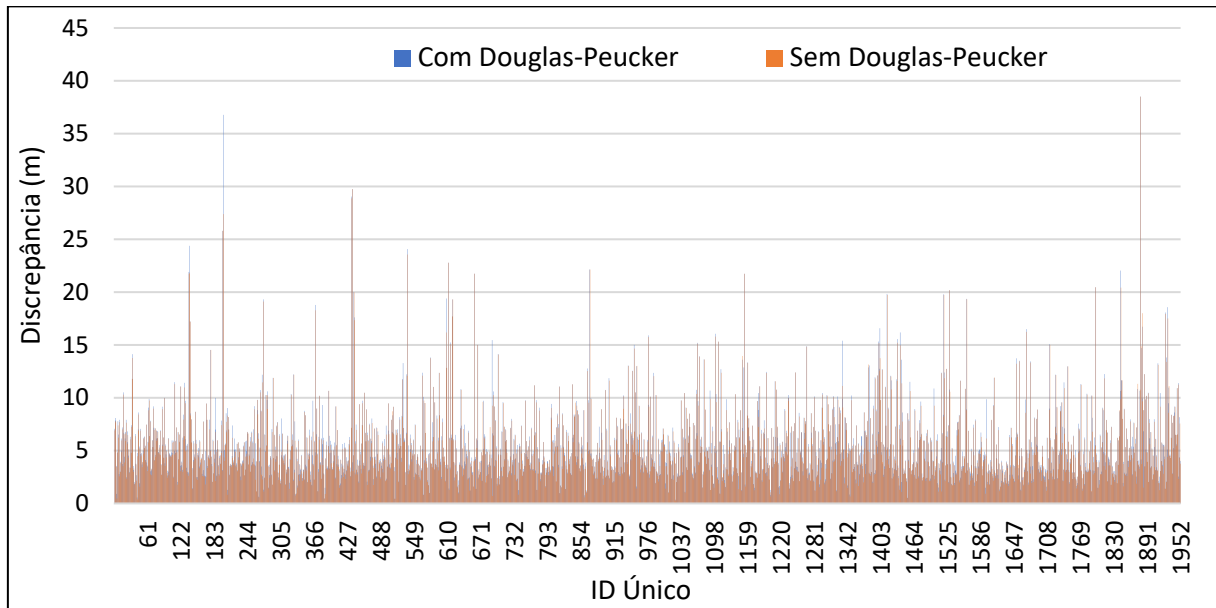
Para isso, foram avaliadas 1952 linhas, correspondentes às estradas da província de Albacete. Estas linhas foram avaliadas quanto a sua acurácia posicional antes e depois de aplicado o algoritmo Douglas-Peucker, utilizando os métodos baseados em análises geométricas. A Figura 14 apresenta os resultados para o método da Distância de Hausdorff.

Figura 14 - Influência do algoritmo Douglas-Peucker no método Distância de Hausdorff.



Já a Figura 15 representa esta análise para a Influência do Vértice.

Figura 15 - Influência do algoritmo Douglas-Peucker no método Influência do Vértice.



De acordo com DSG (2016), para uma escala de 1:100.000, espera-se que pelo menos 90% das discrepâncias de um produto cartográfico fiquem entre 28 m (classe A) e 100 m (classe D). Das feições lineares avaliadas, somente 3 apresentaram discrepância posicional superior a 28 m, no método Influência do Vértice (Figura 15). De fato, a aplicação do algoritmo Douglas-Peucker não produziu diferenças superiores a 5 m nas discrepâncias em 94,5% das linhas para a Distância de Hausdorff e em 99,3% dos casos para a Influência do Vértice (Figura 15).

Desta forma, para esta escala, espera-se que a aplicação do algoritmo Douglas-Peucker não tenha impacto significativo nas análises posteriores estabelecidas a partir das discrepâncias dos métodos baseados em análises geométricas.

De fato, analisando os trabalhos de Santos et al. (2015) e Cunha et al. (2019), pode-se perceber que geralmente os métodos baseados em feições lineares que realizam análises geométricas tendem a apresentar discrepâncias menores que aqueles realizados a partir de faixas de incertezas, como os *buffers*.

Diante disso, foi possível realizar a Simulação Monte Carlo. Durante a simulação, não foi possível obter todos os tamanhos de amostra para as áreas de 1 e 2,5 km<sup>2</sup>. Portanto, para não comprometer as análises, estes tamanhos de área foram excluídos. Foram obtidas amostras até o quinto e o oitavo tamanhos amostrais,

respectivamente. Provavelmente, em uma base de dados de escalar maior seriam obtidos todos os tamanhos de amostra. A Tabela 3 apresenta os tamanhos amostrais analisados, a população utilizada como referência e a área em que a amostra estava inserida (área da janela de busca).

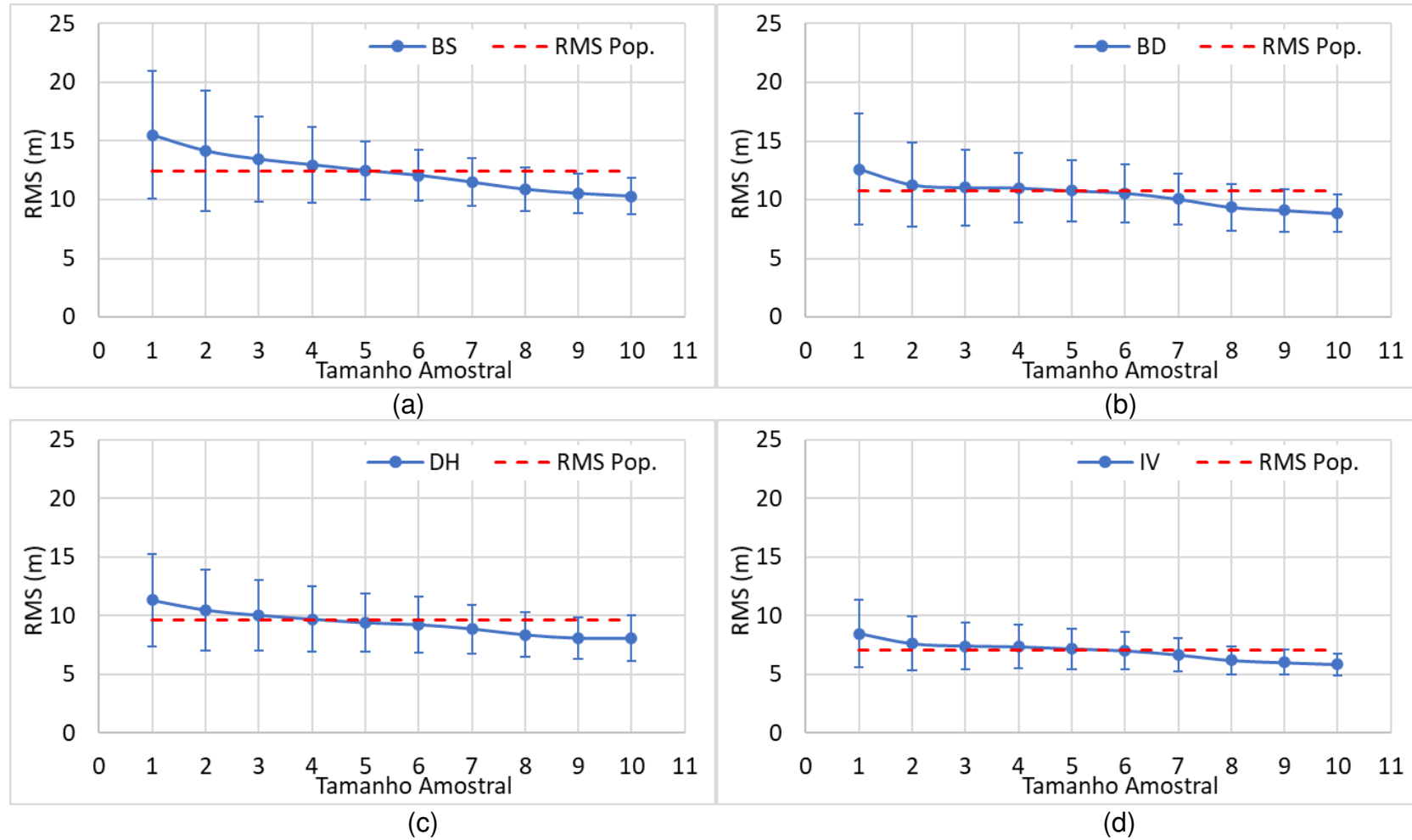
Nesta simulação, vale ressaltar, para cada tamanho amostral foram realizadas 1000 iterações. E para cada iteração foram calculadas as estatísticas das discrepâncias relativas a cada método de feição linear. Por ser uma medida de acurácia, dentre estas estatísticas, foi escolhido o RMS para representar o resultado da simulação. O Apêndice J apresenta os gráficos da relação entre a média do RMS das 1000 iterações e o tamanho amostral, com indicação do respectivo desvio-padrão. A Figura 16 apresenta esta relação para a área de 5 km<sup>2</sup>, como exemplo. De modo geral, pôde-se perceber que o aumento do tamanho da amostra tende a diminuir o desvio-padrão do RMS. Entretanto, com o aumento da área, o desvio-padrão tende a se mostrar uniforme para os diferentes tamanhos amostrais. O único método que não segue este padrão é o *Buffer Duplo*. No Apêndice H (discrepâncias das populações) pode ser observado que, para áreas maiores, surgem valores de discrepância na população que destoam dos demais. Esta pode ser a causa do aumento do desvio-padrão para este método.

Com o aumento da área também se percebeu que, até mesmo para os menores tamanhos amostrais, o valor do RMS amostral se aproximou do valor populacional. Tendo em vista que a proporção entre os comprimentos da amostra e da população foi a mesma e que o RMS populacional pouco variou, isso pode estar relacionado com o número de feições presentes na amostra. Isso porque, para grandes áreas, necessita-se de uma quantidade maior de feições para se obter o comprimento (tamanho amostral) desejado. Esse relativo aumento no número de feições disponíveis pode ter auxiliado as amostras a se aproximarem do valor de RMS (acurácia) da população. Entretanto, para os maiores tamanhos de área, o RMS médio amostral dos métodos *Buffer Duplo* e *Distância de Hausdorff* foi menor que o valor da população. Este não é um comportamento esperado. Entretanto, seria necessário utilizar algum teste estatístico para inferir se esta diferença é significativa.

Tabela 3 - Características das amostras.

| Área da janela de busca (km <sup>2</sup> ) | População (Figura 12) | Tamanho Amostral (km) |        |        |        |        |        |        |        |        |        |
|--|-----------------------|-----------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|  |                       | 1                     | 2      | 3      | 4      | 5      | 6      | 7      | 8      | 9      | 10     |
| 1  | a                     | 1,000                 | 1,500  | 2,000  | 2,500  | 3,000  | 3,500  | 4,000  | 4,500  | 5,000  | 5,500  |
| 2,5  | b                     | 1,581                 | 2,372  | 3,162  | 3,953  | 4,743  | 5,534  | 6,325  | 7,115  | 7,906  | 8,696  |
| 5  | c                     | 2,236                 | 3,354  | 4,472  | 5,590  | 6,708  | 7,826  | 8,944  | 10,062 | 11,180 | 12,298 |
| 7,5  | d                     | 2,739                 | 4,108  | 5,477  | 6,847  | 8,216  | 9,585  | 10,954 | 12,324 | 13,693 | 15,062 |
| 10   | e                     | 3,162                 | 4,743  | 6,325  | 7,906  | 9,487  | 11,068 | 12,649 | 14,230 | 15,811 | 17,393 |
| 15   | f                     | 3,873                 | 5,809  | 7,746  | 9,682  | 11,619 | 13,555 | 15,492 | 17,428 | 19,365 | 21,301 |
| 20   | g                     | 4,472                 | 6,708  | 8,944  | 11,180 | 13,416 | 15,652 | 17,889 | 20,125 | 22,361 | 24,597 |
| 25   | h                     | 5,000                 | 7,500  | 10,000 | 12,500 | 15,000 | 17,500 | 20,000 | 22,500 | 25,000 | 27,500 |
| 50   | i                     | 7,071                 | 10,607 | 14,142 | 17,678 | 21,213 | 24,749 | 28,284 | 31,820 | 35,355 | 38,891 |
| 75   | j                     | 8,660                 | 12,990 | 17,321 | 21,651 | 25,981 | 30,311 | 34,641 | 38,971 | 43,301 | 47,631 |
| 100  | k                     | 10,000                | 15,000 | 20,000 | 25,000 | 30,000 | 35,000 | 40,000 | 45,000 | 50,000 | 55,000 |
| 125  | l                     | 11,180                | 16,771 | 22,361 | 27,951 | 33,541 | 39,131 | 44,721 | 50,312 | 55,902 | 61,492 |
| 150  | m                     | 12,247                | 18,371 | 24,495 | 30,619 | 36,742 | 42,866 | 48,990 | 55,114 | 61,237 | 67,361 |
| 175  | n                     | 13,229                | 19,843 | 26,458 | 33,072 | 39,686 | 46,301 | 52,915 | 59,529 | 66,144 | 72,758 |
| 200  | o                     | 14,142                | 21,213 | 28,284 | 35,355 | 42,426 | 49,497 | 56,569 | 63,640 | 70,711 | 77,782 |
| 225  | p                     | 15,000                | 22,500 | 30,000 | 37,500 | 45,000 | 52,500 | 60,000 | 67,500 | 75,000 | 82,500 |
| 250  | q                     | 15,811                | 23,717 | 31,623 | 39,528 | 47,434 | 55,340 | 63,246 | 71,151 | 79,057 | 86,963 |

Figura 16 – Média do RMS para cada tamanho amostral na área de 5 km<sup>2</sup>. (a) *Buffer Simples*; (b) *Buffer Duplo*; (c) *Distância de Hausdorff* e (d) *Influência do Vértice*.



Em seguida, o teste Kolmogorov-Smirnov foi aplicado para inferir quais tamanhos amostrais poderiam ser considerados estatisticamente iguais à respectiva população. As tabelas a seguir apresentam a estatística **D**, resultado deste teste, para os métodos *Buffer* Simples (Tabela 4), *Buffer* Duplo (Tabela 5), Distância de Hausdorff (Tabela 6) e Influência do Vértice (Tabela 7), além do valor de **D** crítico (**D<sub>c</sub>**).

Cabe destacar que os valores presentes nestas tabelas estão arredondados até a terceira casa decimal e que em negrito estão os valores considerados estatisticamente iguais à população. Para estes valores, que são menores que o valor crítico tabelado do teste Kolmogorov-Smirnov, não se rejeitou a hipótese nula de que as Funções de Distribuição Acumulada das discrepâncias posicionais das amostras e da população são iguais. Como mais de um valor foi considerado igual à população, teve-se que adotar um critério para selecionar o tamanho amostral para representar cada área. Neste estudo, foi escolhido o tamanho amostral mínimo, destacado em vermelho, o que pode ser observado nas Tabelas 4, 5, 6 e 7.

Tabela 4 – Estatística D para o método *Buffer* Simples.

| Área<br>(km <sup>2</sup> ) | Tamanho Amostral |              |              |              |              |              |              |              |              |              | D <sub>c</sub> |       |
|----------------------------|------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|----------------|-------|
|                            | 1                | 2            | 3            | 4            | 5            | 6            | 7            | 8            | 9            | 10           |                |       |
| 5                          | 0,045            | 0,049        | <b>0,023</b> | <b>0,020</b> | <b>0,026</b> | 0,042        | 0,079        | 0,133        | 0,176        | 0,184        | 0,041          |       |
| 7,5                        | 0,051            | 0,062        | 0,059        | <b>0,039</b> | <b>0,018</b> | <b>0,011</b> | <b>0,025</b> | 0,056        | 0,077        | 0,099        | 0,050          |       |
| 10                         | 0,096            | 0,068        | 0,064        | <b>0,052</b> | <b>0,041</b> | <b>0,025</b> | <b>0,011</b> | <b>0,023</b> | <b>0,034</b> | <b>0,046</b> | 0,058          |       |
| 15                         | 0,081            | 0,076        | 0,073        | <b>0,064</b> | <b>0,057</b> | <b>0,048</b> | <b>0,038</b> | <b>0,025</b> | <b>0,020</b> | <b>0,019</b> | 0,071          |       |
| 20                         | 0,080            | 0,093        | 0,082        | 0,072        | 0,069        | <b>0,063</b> | 0,074        | 0,064        | <b>0,062</b> | <b>0,043</b> | <b>0,024</b>   | 0,065 |
| 25                         | 0,098            | 0,085        | 0,081        | 0,079        | <b>0,063</b> | 0,074        | 0,064        | <b>0,062</b> | <b>0,043</b> | <b>0,024</b> | 0,063          | 0,063 |
| 50                         | <b>0,066</b>     | 0,097        | 0,101        | 0,088        | 0,082        | 0,084        | 0,074        | 0,075        | 0,071        | <b>0,063</b> | 0,068          |       |
| 75                         | 0,092            | 0,095        | 0,091        | 0,089        | 0,092        | <b>0,084</b> | <b>0,082</b> | <b>0,074</b> | <b>0,074</b> | <b>0,075</b> | 0,085          |       |
| 100                        | 0,083            | <b>0,074</b> | <b>0,077</b> | 0,089        | 0,083        | 0,084        | 0,082        | 0,087        | <b>0,080</b> | <b>0,079</b> | 0,081          |       |
| 125                        | <b>0,077</b>     | 0,093        | <b>0,080</b> | 0,086        | 0,090        | 0,083        | 0,084        | 0,087        | <b>0,081</b> | <b>0,080</b> | 0,081          |       |
| 150                        | 0,098            | 0,087        | <b>0,080</b> | 0,090        | <b>0,080</b> | <b>0,082</b> | 0,088        | 0,088        | 0,087        | 0,085        | 0,083          |       |
| 175                        | <b>0,079</b>     | 0,089        | 0,086        | 0,082        | <b>0,080</b> | 0,083        | <b>0,078</b> | 0,084        | 0,090        | 0,095        | 0,081          |       |
| 200                        | <b>0,070</b>     | <b>0,080</b> | 0,082        | <b>0,082</b> | <b>0,081</b> | 0,088        | 0,086        | 0,089        | 0,088        | 0,090        | 0,082          |       |
| 225                        | <b>0,070</b>     | <b>0,078</b> | 0,085        | 0,084        | 0,087        | <b>0,079</b> | <b>0,079</b> | 0,084        | <b>0,081</b> | 0,087        | 0,082          |       |
| 250                        | 0,086            | 0,084        | 0,080        | <b>0,074</b> | 0,087        | <b>0,077</b> | 0,080        | <b>0,075</b> | 0,089        | 0,083        | 0,079          |       |

Tabela 5 - Estatística D para o método *Buffer Duplo*.

| Área<br>(km <sup>2</sup> ) | Tamanho Amostral |              |              |              |              |              |              |              |              |              | D <sub>c</sub> |
|----------------------------|------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|----------------|
|                            | 1                | 2            | 3            | 4            | 5            | 6            | 7            | 8            | 9            | 10           |                |
| 5                          | <b>0,025</b>     | <b>0,019</b> | <b>0,013</b> | <b>0,019</b> | <b>0,020</b> | <b>0,014</b> | 0,040        | 0,084        | 0,097        | 0,112        | 0,029          |
| 7,5                        | 0,032            | 0,038        | 0,046        | <b>0,025</b> | <b>0,014</b> | <b>0,027</b> | <b>0,011</b> | 0,032        | 0,050        | 0,060        | 0,029          |
| 10                         | 0,050            | 0,038        | 0,033        | <b>0,031</b> | <b>0,032</b> | <b>0,030</b> | <b>0,025</b> | <b>0,022</b> | <b>0,025</b> | <b>0,030</b> | 0,032          |
| 15                         | 0,043            | 0,040        | 0,042        | <b>0,039</b> | <b>0,031</b> | <b>0,027</b> | <b>0,036</b> | <b>0,030</b> | <b>0,021</b> | <b>0,030</b> | 0,039          |
| 20                         | 0,068            | 0,048        | 0,043        | 0,040        | <b>0,032</b> | <b>0,034</b> | <b>0,031</b> | <b>0,029</b> | <b>0,025</b> | <b>0,019</b> | 0,035          |
| 25                         | 0,066            | 0,071        | 0,047        | <b>0,041</b> | <b>0,038</b> | <b>0,039</b> | <b>0,040</b> | <b>0,043</b> | <b>0,038</b> | <b>0,028</b> | 0,045          |
| 50                         | 0,108            | 0,071        | 0,073        | <b>0,063</b> | <b>0,056</b> | <b>0,050</b> | <b>0,042</b> | <b>0,037</b> | <b>0,038</b> | <b>0,033</b> | 0,064          |
| 75                         | 0,085            | 0,073        | 0,075        | 0,079        | <b>0,061</b> | <b>0,058</b> | <b>0,057</b> | <b>0,042</b> | <b>0,039</b> | <b>0,044</b> | 0,063          |
| 100                        | 0,070            | 0,078        | 0,071        | 0,069        | <b>0,066</b> | <b>0,063</b> | <b>0,049</b> | <b>0,046</b> | <b>0,044</b> | <b>0,041</b> | 0,066          |
| 125                        | 0,073            | 0,086        | 0,085        | <b>0,067</b> | <b>0,065</b> | <b>0,058</b> | <b>0,055</b> | <b>0,053</b> | <b>0,050</b> | <b>0,046</b> | 0,073          |
| 150                        | 0,097            | 0,092        | 0,081        | 0,081        | <b>0,076</b> | <b>0,073</b> | <b>0,063</b> | <b>0,059</b> | <b>0,058</b> | <b>0,048</b> | 0,077          |
| 175                        | 0,099            | 0,097        | <b>0,080</b> | <b>0,083</b> | 0,094        | <b>0,078</b> | <b>0,070</b> | <b>0,070</b> | <b>0,057</b> | <b>0,051</b> | 0,087          |
| 200                        | 0,104            | 0,123        | 0,115        | 0,094        | 0,093        | <b>0,080</b> | <b>0,080</b> | <b>0,078</b> | <b>0,071</b> | <b>0,062</b> | 0,084          |
| 225                        | 0,114            | 0,107        | 0,096        | 0,103        | <b>0,081</b> | <b>0,083</b> | 0,090        | <b>0,062</b> | <b>0,071</b> | <b>0,061</b> | 0,084          |
| 250                        | <b>0,069</b>     | 0,085        | <b>0,079</b> | 0,094        | 0,086        | 0,088        | <b>0,077</b> | <b>0,082</b> | <b>0,081</b> | <b>0,067</b> | 0,084          |

Tabela 6 - Estatística D para o método Distância de Hausdorff.

| Área<br>(km <sup>2</sup> ) | Tamanho Amostral |       |              |              |              |              |              |              |              |              | D <sub>c</sub> |
|----------------------------|------------------|-------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|----------------|
|                            | 1                | 2     | 3            | 4            | 5            | 6            | 7            | 8            | 9            | 10           |                |
| 5                          | <b>0,018</b>     | 0,024 | <b>0,014</b> | <b>0,020</b> | 0,041        | 0,044        | 0,069        | 0,129        | 0,172        | 0,194        | 0,021          |
| 7,5                        | 0,035            | 0,027 | 0,025        | <b>0,018</b> | 0,043        | 0,022        | 0,028        | 0,053        | 0,062        | 0,089        | 0,021          |
| 10                         | 0,035            | 0,039 | 0,029        | <b>0,019</b> | <b>0,016</b> | <b>0,020</b> | 0,027        | 0,028        | 0,028        | 0,035        | 0,021          |
| 15                         | 0,063            | 0,065 | 0,050        | <b>0,032</b> | <b>0,028</b> | <b>0,018</b> | <b>0,021</b> | <b>0,023</b> | <b>0,029</b> | <b>0,030</b> | 0,033          |
| 20                         | 0,088            | 0,069 | 0,060        | <b>0,050</b> | <b>0,034</b> | <b>0,030</b> | <b>0,026</b> | <b>0,022</b> | <b>0,023</b> | <b>0,030</b> | 0,058          |
| 25                         | 0,091            | 0,085 | 0,071        | <b>0,057</b> | 0,062        | <b>0,042</b> | <b>0,032</b> | <b>0,034</b> | <b>0,029</b> | <b>0,025</b> | 0,058          |
| 50                         | 0,149            | 0,107 | 0,102        | <b>0,092</b> | <b>0,083</b> | <b>0,068</b> | <b>0,065</b> | <b>0,059</b> | <b>0,050</b> | <b>0,043</b> | 0,100          |
| 75                         | 0,152            | 0,127 | 0,118        | 0,116        | <b>0,096</b> | <b>0,098</b> | <b>0,089</b> | <b>0,073</b> | <b>0,068</b> | <b>0,060</b> | 0,100          |
| 100                        | 0,156            | 0,150 | 0,149        | 0,129        | <b>0,119</b> | <b>0,115</b> | <b>0,096</b> | <b>0,087</b> | <b>0,080</b> | <b>0,074</b> | 0,129          |
| 125                        | 0,157            | 0,160 | 0,155        | 0,139        | 0,135        | <b>0,124</b> | <b>0,123</b> | <b>0,109</b> | <b>0,098</b> | <b>0,091</b> | 0,129          |
| 150                        | 0,174            | 0,173 | 0,169        | <b>0,151</b> | <b>0,142</b> | <b>0,137</b> | <b>0,125</b> | <b>0,122</b> | <b>0,113</b> | <b>0,094</b> | 0,156          |
| 175                        | 0,182            | 0,184 | 0,164        | 0,163        | 0,162        | <b>0,145</b> | <b>0,137</b> | <b>0,126</b> | <b>0,114</b> | <b>0,100</b> | 0,156          |
| 200                        | 0,179            | 0,195 | 0,189        | 0,171        | 0,170        | <b>0,147</b> | <b>0,145</b> | <b>0,140</b> | <b>0,127</b> | <b>0,109</b> | 0,156          |
| 225                        | 0,195            | 0,199 | 0,177        | 0,185        | <b>0,156</b> | <b>0,152</b> | <b>0,149</b> | <b>0,129</b> | <b>0,126</b> | <b>0,116</b> | 0,165          |
| 250                        | 0,170            | 0,190 | 0,188        | 0,177        | 0,170        | 0,168        | <b>0,142</b> | <b>0,150</b> | <b>0,124</b> | <b>0,115</b> | 0,165          |

Tabela 7 - Estatística D para o método Influência do Vértice.

| Área<br>(km <sup>2</sup> ) | Tamanho Amostral |              |              |              |              |              |              |              |              |              | D <sub>c</sub> |
|----------------------------|------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|----------------|
|                            | 1                | 2            | 3            | 4            | 5            | 6            | 7            | 8            | 9            | 10           |                |
| 5                          | <b>0,020</b>     | <b>0,023</b> | <b>0,013</b> | <b>0,022</b> | <b>0,019</b> | <b>0,019</b> | 0,048        | 0,101        | 0,113        | 0,137        | 0,024          |
| 7,5                        | 0,034            | 0,037        | 0,043        | 0,031        | <b>0,022</b> | <b>0,021</b> | <b>0,010</b> | 0,036        | 0,060        | 0,075        | 0,030          |
| 10                         | 0,061            | 0,041        | 0,035        | <b>0,029</b> | <b>0,030</b> | <b>0,026</b> | <b>0,020</b> | <b>0,018</b> | <b>0,016</b> | <b>0,031</b> | 0,035          |
| 15                         | 0,041            | 0,036        | 0,045        | 0,038        | <b>0,031</b> | <b>0,026</b> | <b>0,033</b> | <b>0,028</b> | <b>0,013</b> | <b>0,021</b> | 0,035          |
| 20                         | 0,054            | 0,053        | 0,045        | 0,039        | <b>0,036</b> | <b>0,036</b> | <b>0,032</b> | <b>0,028</b> | <b>0,012</b> | <b>0,022</b> | 0,039          |
| 25                         | 0,062            | 0,059        | <b>0,047</b> | <b>0,043</b> | <b>0,040</b> | <b>0,042</b> | <b>0,041</b> | <b>0,044</b> | <b>0,034</b> | <b>0,018</b> | 0,055          |
| 50                         | 0,090            | 0,061        | 0,062        | <b>0,053</b> | <b>0,051</b> | <b>0,052</b> | <b>0,045</b> | <b>0,046</b> | <b>0,045</b> | <b>0,039</b> | 0,055          |
| 75                         | 0,077            | 0,064        | 0,071        | 0,074        | 0,063        | <b>0,058</b> | <b>0,056</b> | <b>0,045</b> | <b>0,045</b> | <b>0,047</b> | 0,062          |
| 100                        | 0,074            | 0,080        | 0,077        | 0,080        | 0,077        | <b>0,070</b> | <b>0,054</b> | <b>0,054</b> | <b>0,049</b> | <b>0,047</b> | 0,071          |
| 125                        | 0,074            | 0,082        | 0,078        | <b>0,073</b> | <b>0,067</b> | <b>0,069</b> | <b>0,069</b> | <b>0,061</b> | <b>0,057</b> | <b>0,054</b> | 0,073          |
| 150                        | 0,087            | 0,087        | 0,082        | <b>0,079</b> | <b>0,078</b> | 0,082        | <b>0,069</b> | <b>0,071</b> | <b>0,064</b> | <b>0,056</b> | 0,081          |
| 175                        | 0,101            | 0,093        | 0,089        | <b>0,083</b> | 0,088        | <b>0,079</b> | <b>0,074</b> | <b>0,072</b> | <b>0,064</b> | <b>0,058</b> | 0,086          |
| 200                        | 0,086            | 0,106        | 0,103        | 0,093        | 0,090        | <b>0,079</b> | <b>0,083</b> | <b>0,083</b> | <b>0,079</b> | <b>0,068</b> | 0,085          |
| 225                        | 0,103            | 0,098        | 0,093        | 0,097        | <b>0,079</b> | 0,088        | 0,088        | <b>0,073</b> | <b>0,083</b> | <b>0,071</b> | 0,085          |
| 250                        | <b>0,079</b>     | 0,099        | 0,100        | 0,099        | 0,092        | 0,089        | <b>0,078</b> | 0,085        | <b>0,074</b> | <b>0,071</b> | 0,085          |

Para o método *Buffer* Simples (Tabela 4), não houve nenhuma relação aparente entre o tamanho da área e o tamanho da amostra. Para cinco tamanhos de área (50, 125, 175, 200 e 225 km<sup>2</sup>), o tamanho amostral mínimo foi considerado representativo da população, segundo o teste Kolmogorov-Smirnov. Este fato não é esperado.

Já para os demais métodos, é possível notar um padrão. Observando as Tabelas 5, 6 e 7, percebe-se que na maioria dos casos o quarto ou o quinto tamanho amostral foram os menores comprimentos que foram considerados estatisticamente iguais à população. De fato, dos 60 casos analisados, 36,7% apresentaram o quarto tamanho amostral como o mínimo representativo da população e 21,7%, o quinto.

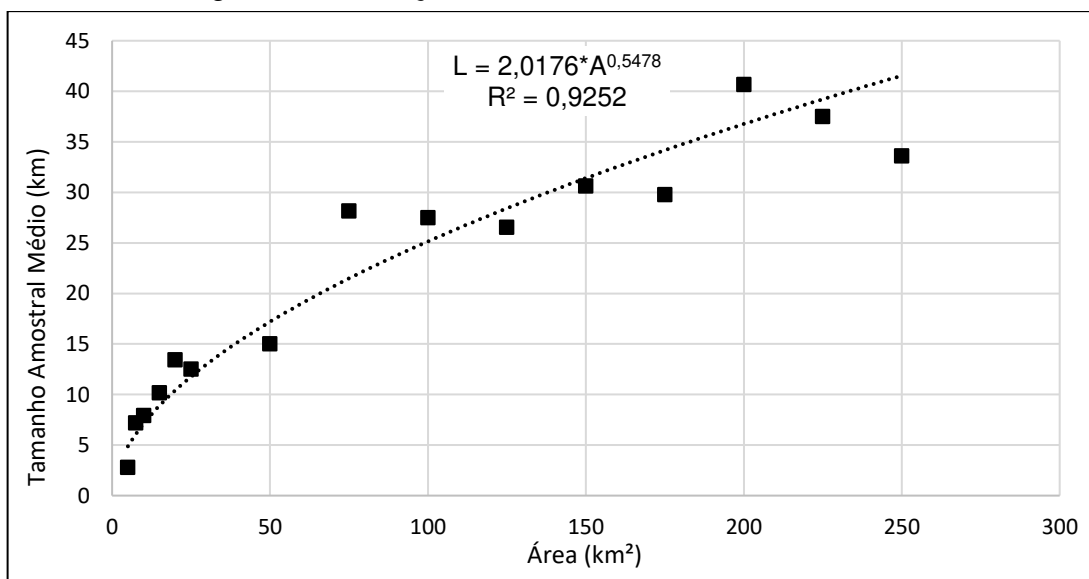
Os tamanhos de amostra mínimos, capazes de representarem a população de discrepâncias, para cada tamanho de área e método de feição linear, estão representados na Tabela 8. Nesta tabela também se encontra a média do tamanho de amostral obtido a partir dos quatro métodos baseados em feições lineares. Cada tipo de método (que utiliza faixa de incerteza ou análises geométricas das linhas) possui potencialidades e limitações. Portanto, um resultado baseado na média representa um equilíbrio entre os métodos, além de fornecer um resultado único, não atrelado a determinado método.

Tabela 8 - Tamanho amostral mínimo representativo da população, por tamanho de área e método.

| Área (km <sup>2</sup> ) | Tamanho Amostral (km) |        |        |        | Média (km) |
|-------------------------|-----------------------|--------|--------|--------|------------|
|                         | BS                    | BD     | DH     | IV     |            |
| 5                       | 4,472                 | 2,236  | 2,236  | 2,236  | 2,795      |
| 7,5                     | 6,847                 | 6,847  | 6,847  | 8,216  | 7,189      |
| 10                      | 7,906                 | 7,906  | 7,906  | 7,906  | 7,906      |
| 15                      | 9,682                 | 9,682  | 9,682  | 11,619 | 10,167     |
| 20                      | 15,652                | 13,416 | 11,180 | 13,416 | 13,416     |
| 25                      | 15,000                | 12,500 | 12,500 | 10,000 | 12,500     |
| 50                      | 7,071                 | 17,678 | 17,678 | 17,678 | 15,026     |
| 75                      | 30,311                | 25,981 | 25,981 | 30,311 | 28,146     |
| 100                     | 15,000                | 30,000 | 30,000 | 35,000 | 27,500     |
| 125                     | 11,180                | 27,951 | 39,131 | 27,951 | 26,553     |
| 150                     | 24,495                | 36,742 | 30,619 | 30,619 | 30,619     |
| 175                     | 13,229                | 26,458 | 46,301 | 33,072 | 29,765     |
| 200                     | 14,142                | 49,497 | 49,497 | 49,497 | 40,659     |
| 225                     | 15,000                | 45,000 | 45,000 | 45,000 | 37,500     |
| 250                     | 39,528                | 15,811 | 63,246 | 15,811 | 33,599     |

A relação entre o tamanho amostral médio e o tamanho da área pode ser observada no gráfico de dispersão presente na Figura 17.

Figura 17 - Relação entre tamanho amostral e área.

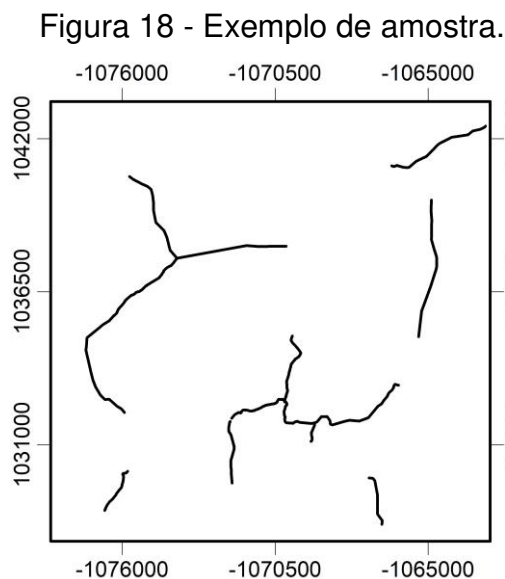


O ajuste de um modelo sobre estes dados indica uma forte associação ( $R^2 = 0,9252$ ) entre o tamanho da área e o tamanho amostral mínimo necessário. O

coeficiente de determinação ( $R^2$ ) indica a proporção de explicação do modelo (BUSSAB e MORETTIN, 2010). Portanto, o modelo obtido neste estudo (Equação 14) pode explicar 92,52% da relação entre o tamanho da área (**A**) e o tamanho amostral (**L**).

$$L = 2,0176 * A^{0,5478} \quad (14)$$

Logo, esta equação, apesar de uma extrapolação, pode ser utilizada para determinação do tamanho amostral mínimo que deve ser utilizado na avaliação da acurácia posicional de produtos cartográficos realizada com feições lineares. A Figura 18 ilustra um exemplo de amostra para uma área de 250 km<sup>2</sup>, de acordo com a Equação 14.



## 6 CONCLUSÃO

O tamanho amostral que deve ser utilizado na avaliação da acurácia posicional realizada com feições lineares é ainda um tema pouco pesquisado. A sua determinação é um fator fundamental para se avançar no controle posicional realizado por meio de linhas.

Este estudo utilizou a Simulação Monte Carlo para gerar amostras de diferentes tamanhos, considerando o padrão de distribuição espacial e o tamanho da área. A

comparação destas amostras com a população foi realizada a partir do teste Kolmogorov-Smirnov.

Uma desvantagem do processo de simulação é o seu alto custo computacional. Mesmo com a utilização do algoritmo Douglas-Peucker, para a exclusão de vértices desnecessários, a simulação demandou 14 computadores processando simultaneamente durante seis dias ininterruptos. Entretanto, apesar de ser um processo complexo, a Simulação Monte Carlo forneceu uma base estatística e confiabilidade na análise realizada sobre as amostras, possibilitando a avaliação de diversos cenários.

Por fim, este estudo apresentou uma proposta para a determinação do tamanho amostral de feições lineares utilizadas no controle posicional, em forma de equação. Além desta, também é indicada uma versão simplificada, ambas relacionadas com o tamanho da área avaliada.

Como conclusão geral, diante do grande número de amostras simuladas e dos resultados, como o valor coeficiente de determinação entre o tamanho da área e o tamanho amostral, pode-se dizer que a proposta apresentada é coerente e fornece uma contribuição real para o tema, que ainda é pouco pesquisado.

Para trabalho futuros, recomenda-se replicar esta metodologia para outros tipos de feições, como pontos ou mesmo para linhas tridimensionais.

## **REFERÊNCIAS BIBLIOGRÁFICAS**

**ABBAS, I. Base de Données Vectorielles et Erreur Cartographique. Problèmes Posés Par Le Controle Ponctuel. Une Méthode Alternative Fondée Sur La Distance de Hausdorff: Le Controle Linéaire.** Thèse de Doctorat en Informatique. Université Denis–Diderot, Paris, France, 1994.

**AENOR - ASOCIACIÓN ESPAÑOLA DE NORMALIZACIÓN Y CERTIFICACIÓN. UNE 148002:2016: Metodología de evaluación de la exactitud posicional de la información geográfica.** Madrid. 46p. 2016.

**ASPRS - AMERICAN SOCIETY FOR PHOTOGRAMMETRY AND REMOTE SENSING. ASPRS 2014: Positional Accuracy Standards for Digital Geospatial Data.**

**Photogrammetric Engineering & Remote Sensing**, v.81, n.3, pA1-A26. Disponível em [https://www.asprs.org/wp-content/uploads/2015/01/ASPRS\\_Positional\\_Accuracy\\_Standards\\_Edition1\\_Version100\\_November2014.pdf](https://www.asprs.org/wp-content/uploads/2015/01/ASPRS_Positional_Accuracy_Standards_Edition1_Version100_November2014.pdf)>. Acesso em 20 de mar. de 2023. 2015.

ARIZA-LÓPEZ, F. J. **Calidad en la producción cartográfica**. Madrid: Ra–Ma, 389 p. 2002a.

ARIZA-LÓPEZ, F.J. **Trabajo de investigación: Curva Operativa para el control de Calidad Posicional en Cartografía**. Concurso para la provisión de una plaza de Catedrático de Universidad, Departamento de Ingeniería Cartográfica, Geodesia y Fotogrametría. Jun., Universidad de Jaén, Jaén, Espanha, 2002b.

ARIZA-LÓPEZ, F. J. L.; MOZAS-CALVACHE, A. T. Methodology for the Development of a Sampling Recommendation When Using Linear Elements for Positional Quality Control. In: **International Cartographic Conference**. Espanha, 2005.

ARIZA-LÓPEZ, F. J., ATKINSON-GORDO, A. D. J.; NERO, M. A. Análisis de Algunas Metodologías de Evaluación de la Componente Posicional. **Topografía y cartografía: Revista del Ilustre Colegio Oficial de Ingenieros Técnicos en Topografía**, v. 24, p. 32-45, 2007.

ARIZA-LÓPEZ, F. J., ATKINSON-GORDO, A. D. J., NERO, M. A.; CINTRA, J. P. La Componente Posicional de los Datos Geográficos: análisis crítico de los Métodos de Evaluación y Reporte en Hispanoamérica. **Revista Cartográfica**, v. 83, p. 95-152, 2007.

ARIZA-LÓPEZ, Francisco Javier; ATKINSON-GORDO, Alan David. Analysis of some positional accuracy assessment methodologies. **Journal of surveying Engineering**, v. 134, n. 2, p. 45-54, 2008.

ARIZA-LÓPEZ, F. J., MOZAS-CALVACHE, A. T., UREÑA-CÁMARA, M. A., ALBA-FERNÁNDEZ, V., GARCÍA-BALBOA, J. L., RODRÍGUEZ-AVI, J., RUIZ-LENDÍNEZ, J. J. Influence of sample size on line-based positional assessment methods for road data. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 66, n. 5, p. 708-719, 2011. DOI: 10.1016/j.isprsjprs.2011.06.003.

ARIZA-LÓPEZ, Francisco Javier; MOZAS-CALVACHE, Antonio Tomás. Comparison of four line-based positional assessment methods by means of synthetic data. **Geoinformatica**, v. 16, n. 2, p. 221-243, 2012.

ARIZA-LÓPEZ, Francisco; RUIZ-LENDÍNEZ, Juan; UREÑA-CÁMARA, Manuel. Influence of Sample Size on Automatic Positional Accuracy Assessment Methods for Urban Areas. **ISPRS International Journal of Geo-Information**, v. 7, n. 6, p. 200, 2018.

BRASIL. Decreto N° 89.817 de 20 de junho de 1984. **Estabelece as Instruções Reguladoras das Normas Técnicas da Cartografia Nacional**. Brasil. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/decreto/1980-1989/D89817.htm](http://www.planalto.gov.br/ccivil_03/decreto/1980-1989/D89817.htm)>.

BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica**. 6. ed. São Paulo: Editora Saraiva, 2010.

BUSLENKO, Nikolaï Panteleïmonovich. **The Monte Carlo Method**: the method of statistical trials. Pergamon Press, 1966.

CINTRA, J. P.; NERO, M. A. New method for positional cartographic quality control in digital mapping. **Journal of Surveying Engineering**, v. 141, p. 04015001-1-04015001-10, 2015. [https://doi.org/10.1061/\(ASCE\)SU.1943-5428.0000141](https://doi.org/10.1061/(ASCE)SU.1943-5428.0000141).

CUNHA, M. M.; SECATTO, G. Z.; GALINDO, J. R. F.; SANTOS, A. P. Proposta de um método de avaliação da acurácia posicional baseado na modificação do Buffer Simples. **Revista Brasileira de Cartografia**, v. 71, n. 4, p. 1193-1218, 2019.

DOUGLAS, D. H.; PEUCKER, T. K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. **The Canadian Cartographer**, v. 10, n. 2, p. 112-122, 1973.

DSG - DIRETORIA DO SERVIÇO GEOGRÁFICO. **Especificação Técnica Para Controle de Qualidade de Dados Geoespaciais (ET-CQDG)**. Brasília, 2016.

FGDC - FEDERAL GEOGRAPHIC DATA COMMITTEE. **Geospatial Positioning Accuracy Standards Part 3: National Standard for Spatial Data Accuracy**. United States. 1998.

FONSECA, I. G. R. **Avaliação da Acurácia Posicional Tridimensional Utilizando Feições Lineares**. Dissertação (Mestrado em Engenharia Civil). Viçosa: Universidade Federal de Viçosa. 84p. 2018.

FREEDMAN, D.; DIACONIS, P. On the histogram as a density estimator: L<sup>2</sup> theory. *Zeitschr. Wahrsch. Theor. Verwan. Gebiete* 57(4), pp. 453–476, 1981.

GOODCHILD, M. F.; HUNTER, G. J. A simple positional accuracy measure for linear features, **International Journal of Geographical Information Science**, 11:3, 299-306, 1997, DOI: 10.1080/136588197242419.

GUELFÍ, Marcelo; LÓPEZ-VÁZQUEZ, Carlos. Comparación del método de Thiessen con alternativas más simples mediante simulación de Monte Carlo. **Revista Cartográfica**, n. 91, p. 143-157, 2015.

HARDIN, Perry J.; SHUMWAY, J. Matthew. Statistical significance and normalized confusion matrices. **Photogrammetric Engineering and Remote Sensing**, v. 63, n. 6, p. 735-739, 1997.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Manuais Técnicos em Geociências. **Avaliação da Qualidade de Dados Geoespaciais**. 2 ed. Rio de Janeiro, RJ, 117 p, 2019.

ICSM - INTERGOVERNMENTAL COMMITTEE ON SURVEY AND MAPPING. **Australian Map and Spatial Horizontal Data Accuracy Standard**. Australia. Disponível em: <<https://www.icsm.gov.au/publications/spatial-data-horizontal-accuracy-standard>>. Acesso em: 01 de mar. de 2023. 2009.

INSTITUTO GEOGRÁFICO NACIONAL. **Mapa Topográfico Nacional 1:25.000: MTN25**. Madrid, Espanha: IGN, 2021a. Shapefile. Escala 1:25.000. Disponível em: <http://centrodedescargas.cnig.es/CentroDescargas/buscadorCatalogo.do>. Acesso em: 1 dez. 2022.

INSTITUTO GEOGRÁFICO NACIONAL. **Base Topográfica Nacional 1:100.000: BTN100**. Madrid, Espanha: IGN, 2021b. Shapefile. Escala 1:100.000. Disponível em: <http://centrodedescargas.cnig.es/CentroDescargas/buscadorCatalogo.do>. Acesso em: 1 dez. 2022.

KALOS, M. H.; WHITLOCK, P. A. **Monte Carlo Methods**. John Wiley & Sons, 2009.

LA VEGA, P. G; ARIZA-LÓPEZ, F. J.; MOZAS-CALVACHE, A. T. Models for positional accuracy assessment of linear features: 2D and 3D cases. **Survey Review**, v. 48, n. 350, p. 347-360, 2016.

MOREIRA, J. M.; OLIVEIRA, G. A.; SILVA, P. L.; SANTOS, A. P.; MARQUES, E. T. Avaliação Da Qualidade Posicional Planimétrica De Ortofotos Usadas No Cadastro Territorial Multifinalitário Através De Feições Lineares. In: CONGRESSO DE CADASTRO MULTIFINALITÁRIO E GESTÃO TERRITORIAL, **Anais...**, Florianópolis, SC, 2018.

MOZAS-CALVACHE, A. T.; ARIZA-LÓPEZ, F. J. Methodology for positional quality control in cartography using linear features. **The Cartographic Journal**, v. 47, n. 4, p. 371-378, 2010.

MOZAS-CALVACHE, A. T. **Control de calidad posicional en cartografía por elementos lineales**. Tese (Doutorado em Ingeniería Cartográfica, Geodésica y Fotogrametría). Jaén: Universidad de Jaén, Espanha. 249p. 2007.

MOZAS-CALVACHE, A. T.; ARIZA-LÓPEZ, F. J. New method for positional quality control in cartography based on lines. A comparative study of methodologies. **International Journal of Geographical Information Science**, v. 25, n. 10, p. 1681-1695. 2011. DOI:10.1080/13658816.2010.545063.

NERO, Marcelo Antônio. **Propostas para o controle de qualidade de bases cartográficas com ênfase na componente posicional**. Tese (Doutorado em Engenharia). Universidade de São Paulo. 2005.

NERO, M. A.; CINTRA, J. P.; FERREIRA, G. F.; PEREIRA, T. A. J.; FARIA, T. S. A computational tool to evaluate the sample size in map positional accuracy. **Boletim de Ciências Geodésicas [online]**, vol.23, n.3, pp.445-460, 2017. <https://doi.org/10.1590/s1982-21702017000300030>.

NOGUEIRA JÚNIOR, João Bosco. **Controle de qualidade de produtos cartográficos: uma proposta metodológica**. Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de Ciências e Tecnologia, 131 p, 2003.

NOGUEIRA JÚNIOR, J. B.; MONICO, J. F. G.; TACHIBANA, V. M. Tamanho da amostra no controle de qualidade posicional de dados cartográficos. **Boletim de Ciências Geodésicas [online]**, v. 10, n.1, p. 101-112, 2004.

NSA - NATO STANDARDIZATION AGENCY. **STANAG (Standardization Agreement) 2215: Evaluation of land maps, aeronautical charts and digital topographic data**. North Atlantic Treaty Organization (NATO). Bruxelas, 7ª Ed, 2010.

ORTIZ, Jussara Oliveira; FELGUEIRAS, Carlos Alberto; RENNÓ, Camilo Daleles. Simulação conjunta de variáveis correlacionadas para aplicação em modelagem espacial. In: XIII Simpósio Brasileiro de Sensoriamento Remoto, **Anais...**, Florianópolis, SC, 2007.

PENHA, Alex de Lima Teodoro; COELHO, Vagner Braga Nunes. Reconhecimento da existência de um desvio padrão máximo a partir de perturbação de pontos para o qual a topologia de redes triangulares irregulares permanece constante. In: XVII Simpósio Brasileiro de Sensoriamento Remoto - SBSR, **Anais...**, João Pessoa, PB, 2015.

PEREIRA, T. A. J.; NERO, M. A. Análise de normas de controle de qualidade posicional em Cartografia ao redor do Mundo: exemplos atuais de alguns países. In: IV SIMGEO - IV Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação, 2012, Recife-PE. **Anais...** Recife-PE: Editora UFPE, v. 1. p. 001-0010, 2012.

R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<https://www.R-project.org/>>. Acessado em: 08 de fevereiro de 2023. 2023.

REIS, M. S.; DUTRA, L. V.; SANT'ANNA, S. J. S.; ESCADA, M. I. S. Análise das incertezas envolvidas em classificação multi-legendas da cobertura da terra com suporte de simulação Monte Carlo. **Revista Brasileira de Cartografia**, v. 69, n. 9, 2017.

RÍOS, D.; RÍOS, S.; MARTÍN, J. **Simulación, Métodos y Aplicaciones**. Ra-Ma: Madrid, Spain, 1997.

ROBERT, Christian; CASELLA, George. **Monte Carlo Statistical Methods**. Springer Science & Business Media, 2013.

ROFATTO, V. F.; MATSUOKA, M. T.; KLEIN, I. Método Monte Carlo Aplicado À Análise De Confiabilidade Em Geodésia. In: XXVII CONGRESSO BRASILEIRO DE CARTOGRAFIA E XXVI EXPOCART, **Anais...**, - Rio de Janeiro, RJ, 2017.

ROSS, S. M. **Simulation**. 4th edition, San Diego: Academic Press, 2006.

SANTOS, A. P., MEDEIROS, N. G., SANTOS, G. R., RODRIGUES, D. D. Controle de qualidade posicional em dados espaciais utilizando feições lineares. **Boletim de Ciências Geodésicas**, v. 21, n. 2, p. 233-250, 2015.

SANTOS, A. P. **Controle De Qualidade Cartográfica: Metodologias Para Avaliação Da Acurácia Posicional Em Dados Espaciais**. Tese (Doutorado em Engenharia Civil) – Universidade Federal de Viçosa, Viçosa, 188p, 2015.

SANTOS, A. P., MEDEIROS, N. G., SANTOS, G. R. D., RODRIGUES, D. D. Avaliação da acurácia posicional planimétrica em modelos digitais de superfície com o uso de feições lineares. **Boletim de Ciências Geodésicas**, v. 22, n. 1, p. 157-174. 2016.

SCOTT, D. W. On optimal and data-based histograms. **Biometrika**, v. 66, n. 3, p. 605-610, 1979.

SILVA, Daniel Carneiro; DALMOLIN, Quintino. Mapeamento incremental: solução teórica para fotogrametria em áreas nubladas. **Boletim de Ciências Geodésicas**, v. 8, n. 2, 2002.

SILVA, Larissa Antunes. **Comparação de características visíveis de relâmpagos nuvensolo negativos para diferentes dias de tempestade no sudeste do Brasil, observados por uma rede de câmeras de alta velocidade**. Dissertação (Mestrado em Geofísica Espacial/Ciências Atmosféricas) – Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2014.

STURGES, H. A. The Choice of a Class Interval. **Journal of the American Statistical Association**, 21(153), 65–66, 1926.

TVEITE, Havard; LANGAAS, Sindre. An accuracy assessment method for geographical line data sets based on buffering, **International Journal of Geographical Information Science**, v. 13, n. 1, p. 27-47, 1999, DOI: 10.1080/136588199241445.

WAND, M. P. Data-based choice of histogram bin width. **The American Statistician**, v. 51, n. 1, p. 59-64, 1997.

ZANETTI, J., PAULA, R. M.; SANTOS, A. P.; MEDEIROS, N. G. Avaliação da acurácia posicional planimétrica de ortoimagens disponibilizadas nos sistemas de informações geográficas. **Revista Brasileira de Cartografia**, v. 68, n. 7, 2016.

## CAPÍTULO 3. PROPOSTA DE UMA METODOLOGIA PARA AVALIAÇÃO DA ACURÁCIA POSICIONAL UTILIZANDO FEIÇÕES LINEARES: O CASO DA NORMA BRASILEIRA E EXEMPLOS DE APLICAÇÃO

### Resumo

A qualidade da informação é um fator preponderante em qualquer área do conhecimento. Nas informações espaciais a qualidade de um produto cartográfico é avaliada segundo seus elementos básicos: completude, consistência lógica, acurácia posicional, acurácia temporal, acurácia temática e usabilidade. Dentre estes, a acurácia posicional é o mais utilizado para retratar a qualidade dos dados espaciais. Para a sua avaliação pelo método tradicional por pontos existem padrões, normas e metodologias já estabelecidas, o que não ocorre quando a avaliação da acurácia posicional é realizada com feições lineares. A análise da qualidade posicional de um dado espacial por meio de linhas, sem uma metodologia definida, impede a sua integração em uma base de dados e pode prejudicar a tomada de decisões realizadas com auxílio deste dado espacial, por exemplo. Diante disso, este estudo propôs uma metodologia geral para a avaliação da acurácia posicional planimétrica realizada com feições lineares, segundo a norma brasileira. Esta metodologia compreende: determinação do tamanho amostral, análise do padrão de distribuição espacial, detecção de *outlier*, e análises de tendência e de precisão. Esta metodologia foi utilizada em três bases cartográficas: de pequena, média e grande escala e os resultados comparados com a avaliação resultante do método tradicional, realizado com pontos. Para todas as escalas, a metodologia proposta apresentou resultado compatível, ou mais rigoroso, que o método de feições pontuais, demonstrando a viabilidade de sua utilização.

### 1 INTRODUÇÃO

Em todas as áreas do conhecimento a qualidade da informação utilizada é um dos fatores determinantes para a eficácia do processo (SHI, FISHER e GOODCHILD, 2005). Segundo ISO (2005), esta qualidade pode ser entendida como o grau de adequação das características inerentes de um produto a determinados requisitos. Geralmente, estes requisitos são organizados em forma de um padrão, norma ou metodologia.

Nas informações espaciais a qualidade dos produtos cartográficos é avaliada segundo os elementos básicos: a acurácia posicional, a acurácia temática, a acurácia temporal, a consistência lógica, a completude e a usabilidade, descritos na ISO19157 (ISO, 2013). Ariza-López et al. (2021) afirmam que a avaliação da qualidade de dados

espaciais segundo estes elementos deve ser baseada em métodos padronizados e bem definidos.

Para o caso da acurácia posicional, no Brasil, desde 1984, o Decreto nº 89.817 regulamenta a avaliação posicional de produtos cartográficos (BRASIL, 1984). Posteriormente, este documento foi complementado pela Especificação Técnica para Controle de Qualidade de Dados Geoespaciais (DSG, 2016a). Além de incluir uma classe mais restritiva para analisar a qualidade posicional, tendo em vista o avanço da Cartografia Digital, este texto trouxe metodologias para a avaliação dos demais elementos de qualidade. Baseado nestas normas, o trabalho Santos et al. (2016) apresenta uma metodologia para a avaliação da acurácia posicional por pontos muito difundida no Brasil.

Entretanto, as metodologias para avaliação de precisão posicional não são estáticas, elas evoluem com o tempo (ARIZA-LÓPEZ et al., 2021). Mozas-Calvache e Ariza-López (2010) afirmam que uma evolução dos métodos baseados em feições pontuais seria a utilização de feições lineares no controle da qualidade posicional em Cartografia.

De fato, nas últimas décadas diversos métodos baseados em feições lineares foram propostos. Dentre os mais utilizados estão o *Buffer* Simples (GOODCHILD e HUNTER, 1997), o *Buffer* Duplo (TVEITE e LANGAAS, 1999), a Influência do Vértice (MOZAS-CALVACHE e ARIZA-LÓPEZ, 2011) e a Distância de Hausdorff (ABBAS, 1994). Entretanto, estes métodos são utilizados ainda de forma incipiente no Brasil e, quando utilizados, são aplicados sem nenhuma padronização. No Brasil, o único estudo que se tem conhecimento que propõe uma metodologia para utilizar as discrepâncias provenientes de feições lineares no controle posicional, segundo a norma brasileira, é apresentado por Santos et al. (2015). Entretanto, estes autores não apresentam como deve ser determinado o tamanho da amostra de controle e sua distribuição espacial.

Por se tratar de um processo baseado em amostragem, é de se esperar que o tamanho da amostra possa interferir no resultado da avaliação. Dessa forma, a ausência de um método para determinar o tamanho da amostra impede a padronização do processo de avaliação da acurácia posicional realizada com feições lineares. Por sua vez, por afetar a sua representatividade perante à população, a

distribuição espacial da amostra pode determinar a validade do processo de avaliação (ARIZA-LÓPEZ e ATKINSON-GORDO, 2008).

A falta de uma metodologia que inclua a padronização destes aspectos pode, por exemplo, impedir ou dificultar a integração de produtos cartográficos avaliados utilizando-se feições lineares em uma mesma base cartográfica. Isso porque não há, até então, uma homogeneização dos parâmetros utilizados na avaliação dos produtos.

A falta de uma metodologia única também pode levar a tomadas de decisões equivocadas, fazendo, por exemplo, com que o analista rejeite um bom produto (risco do produtor) ou aceite um produto não adequado ao propósito a que se destina (risco do usuário).

Diante do exposto, o objetivo deste capítulo é propor uma metodologia geral, bem definida e estatisticamente embasada para o controle posicional, baseada em feições lineares, que compreenda a padronização do tamanho, a distribuição espacial da amostra, a detecção de *outliers* e as análises de precisão e tendência.

## **2 NORMAS BRASILEIRAS PARA A AVALIAÇÃO DA ACURÁCIA POSICIONAL**

No Brasil existem duas normas principais que abordam a qualidade no âmbito da Cartografia: o Decreto nº 89.817/84 (BRASIL, 1984) e a Especificação Técnica para Controle de Qualidade de Dados Geoespaciais (ET-CQDG) da Infraestrutura Nacional de Dados Espaciais (INDE) (DSG, 2016a).

O Decreto nº 89.817 de 20 de junho de 1984 (BRASIL, 1984) é a norma que estabelece as instruções reguladoras da Cartografia Nacional. Este documento estabelece as normas e padrões mínimos a serem seguidos por entidades públicas e privadas produtoras e usuárias de produtos cartográficos (BRASIL, 1984).

Este decreto possui duas seções destinadas ao controle de qualidade cartográfica. Em uma destas seções é apresentado o Padrão de Exatidão Cartográfica (PEC), um indicador estatístico de dispersão, relativo a 90% de probabilidade, utilizado como medida para avaliar a acurácia posicional de produtos cartográficos (BRASIL, 1984). Outra medida utilizada por este decreto é o Erro-Padrão (EP) que, para esta norma, é sinônimo de desvio-padrão. A relação entre o PEC e o EP é dada pela Equação 1.

$$PEC = 1,6449 \times EP \quad (1)$$

A Tabela 1 mostra os valores de PEC e EP, para a componente planimétrica, utilizados para enquadrar os produtos cartográficos segundo as classes presentes no Decreto nº 89.817.

Tabela 1 - Valores de PEC e EP para a planimetria segundo o Decreto nº89.817/84.

| Classe | Planimetria  |              |
|--------|--------------|--------------|
|        | PEC          | EP           |
| A      | 0,50 mm x DE | 0,30 mm x DE |
| B      | 0,80 mm x DE | 0,50 mm x DE |
| C      | 1,00 mm x DE | 0,60 mm x DE |

DE = denominador da escala

Fonte: Brasil (1984).

DSG (2016a) afirma que estas tolerâncias refletem os métodos e técnicas disponíveis na época da criação desse decreto.

O surgimento de novas geotecnologias contribuiu para que fosse realizada uma revisão dos padrões estabelecidos no Decreto nº 89.817. Em consequência disso, a Especificação Técnica para Estruturação dos Dados Geoespaciais Vetoriais (ET-ADGV), da INDE, estabeleceu um novo Padrão de Exatidão Cartográfica destinado a atender os Produtos Cartográficos Digitais (PEC-PCD) (DSG, 2011). Entretanto, vislumbrando a necessidade de um documento específico para o controle de qualidade cartográfica, em 2016 a INDE publicou a ET-CQDG, que é baseada na norma ISO 19157:2013 (ISO, 2013).

Esse novo padrão, apresentado pela ET-ADGV e incorporado pela ET-CQDG, fez com que surgisse uma classe mais restritiva do que aquelas apresentadas no decreto. A Tabela 2 apresenta as classes da ET-CQDG (DSG, 2016a) e do Decreto nº89.817 (BRASIL, 1984).

Tabela 2 - Valores de PEC e EP para a planimetria.

| Classes             |         | Planimetria  |              |
|---------------------|---------|--------------|--------------|
| Decreto nº89.817/84 | ET-CQDG | PEC-PCD      | EP           |
| -                   | A       | 0,28 mm x DE | 0,17 mm x DE |
| A                   | B       | 0,50 mm x DE | 0,30 mm x DE |
| B                   | C       | 0,80 mm x DE | 0,50 mm x DE |
| C                   | D       | 1,00 mm x DE | 0,60 mm x DE |

DE = denominador da escala

Fonte: DSG (2016a).

DSG (2016a) destaca que, apesar do acréscimo da classe mais restritiva, o procedimento de cálculo para os valores de **PEC** e **EP** são os mesmos do decreto.

Apesar de ser uma norma muito importante para a Cartografia brasileira, Santos et al. (2016) mostra que o Decreto nº89.817 apresenta alguns erros conceituais que podem comprometer sua aplicação. Neste sentido, a ET-CQDG pode ser utilizada para complementar e ajudar a aplicar a avaliação da acurácia posicional proposta pelo Decreto nº 89.817.

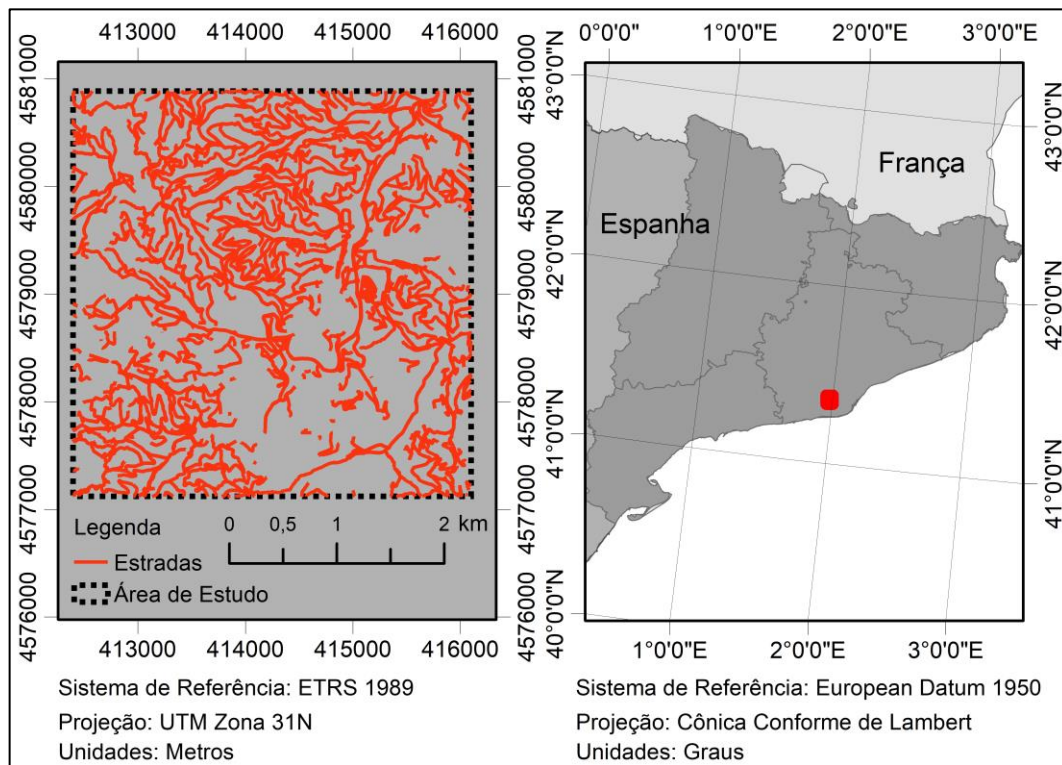
### 3 MATERIAIS E MÉTODOS

#### 3.1 Base de Dados

A primeira base de dados utilizada nesta pesquisa foi obtida no Instituto Cartográfico e Geológico da Catalunha (ICGC) da Espanha. Desta base foram selecionados 407 km de estradas, presentes em uma área de aproximadamente 14 km<sup>2</sup> no sul da província de Barcelona (Figura 1).

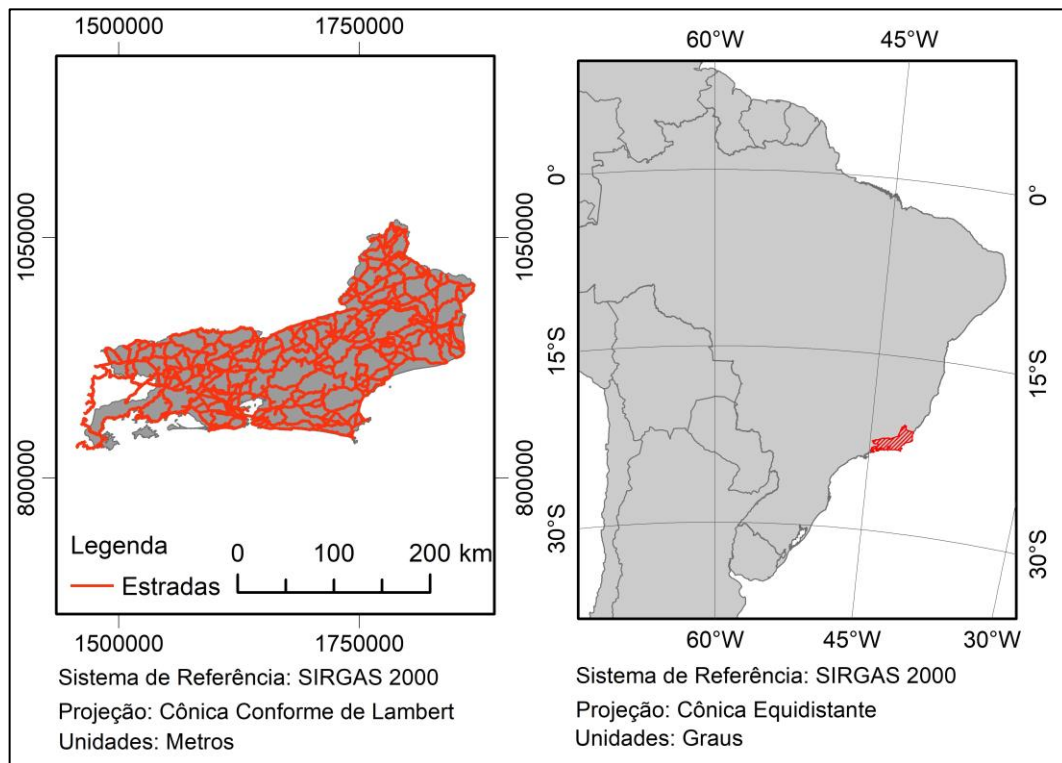
As feições lineares de referência são provenientes da Cartografia Topográfica na escala de 1/1.000 (INSTITUTO CARTOGRÁFICO E GEOLÓGICO DA CATALUNHA, 2021a). Já as feições lineares de teste foram obtidas na Base Topográfica na escala de 1/5.000 (INSTITUTO CARTOGRÁFICO E GEOLÓGICO DA CATALUNHA, 2021b). Estes dados possuem como sistema de referência o ETRS (*European Terrestrial Reference System*) 1989 e sua projeção é a UTM, zona 31N.

Figura 1 – Parte da base de dados da Catalunha.



Já a segunda base de dados utilizada neste trabalho é composta pelas estradas da base cartográfica contínua do Rio de Janeiro, que cobre uma área de 43750 km<sup>2</sup>. Foram utilizadas as feições na escala 1/25.000 como referência para se avaliar as linhas homólogas na escala 1/250.000 (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2023a, 2023b). Nesta área ao todo são 8425 km de estradas, como pode ser observado na Figura 2. Estes dados são disponibilizados como bases cartográficas contínuas pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Estes dados possuem como sistema de referência o SIRGAS 2000 e sua projeção é a Cônica Conforme de Lambert.

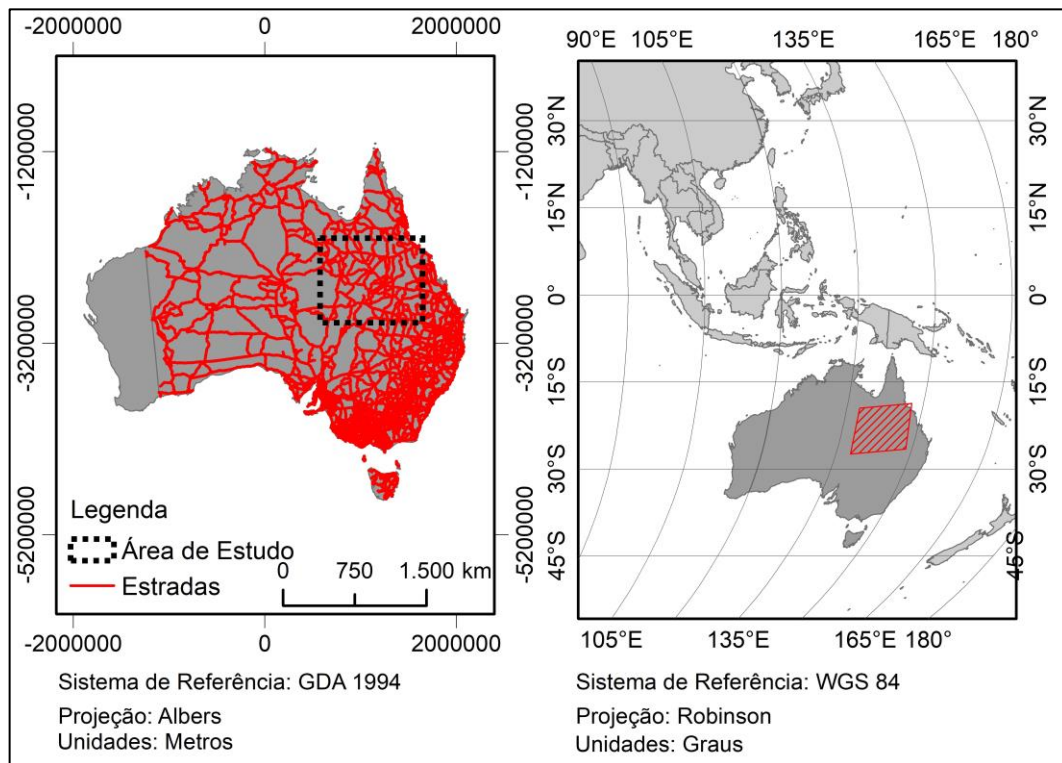
Figura 2 - Base cartográfica contínua do Rio de Janeiro.



A terceira base de dados (Figura 3) compreende 18150 km de estradas do nordeste da Austrália, presentes em uma área de 944530 km<sup>2</sup>, e fornecidos pela Geoscience Australia. Os dados de referência correspondem às linhas na escala de 1/1.000.000, que foram utilizados para avaliar as feições de teste homólogas na escala de 1/5.000.000 (GEOSCIENCE AUSTRALIA, 2001, 2004). Estes dados possuem como sistema de referência o GDA (*Geocentric Datum of Australia*) 1994 e utilizam a projeção de Albers.

Em um processo de avaliação da qualidade posicional é sempre importante saber se a base de teste é derivada da base de referência, pois isso pode afetar a qualidade dos dados de teste. Na área do Rio de Janeiro, as feições de referência são da base cartográfica contínua, enquanto as feições de teste são provenientes do mapeamento sistemático. Portanto, são mapeamentos diferentes. Já para as áreas da Catalunha e da Austrália, esta informação não estava disponível.

Figura 3 - Estradas da Austrália.

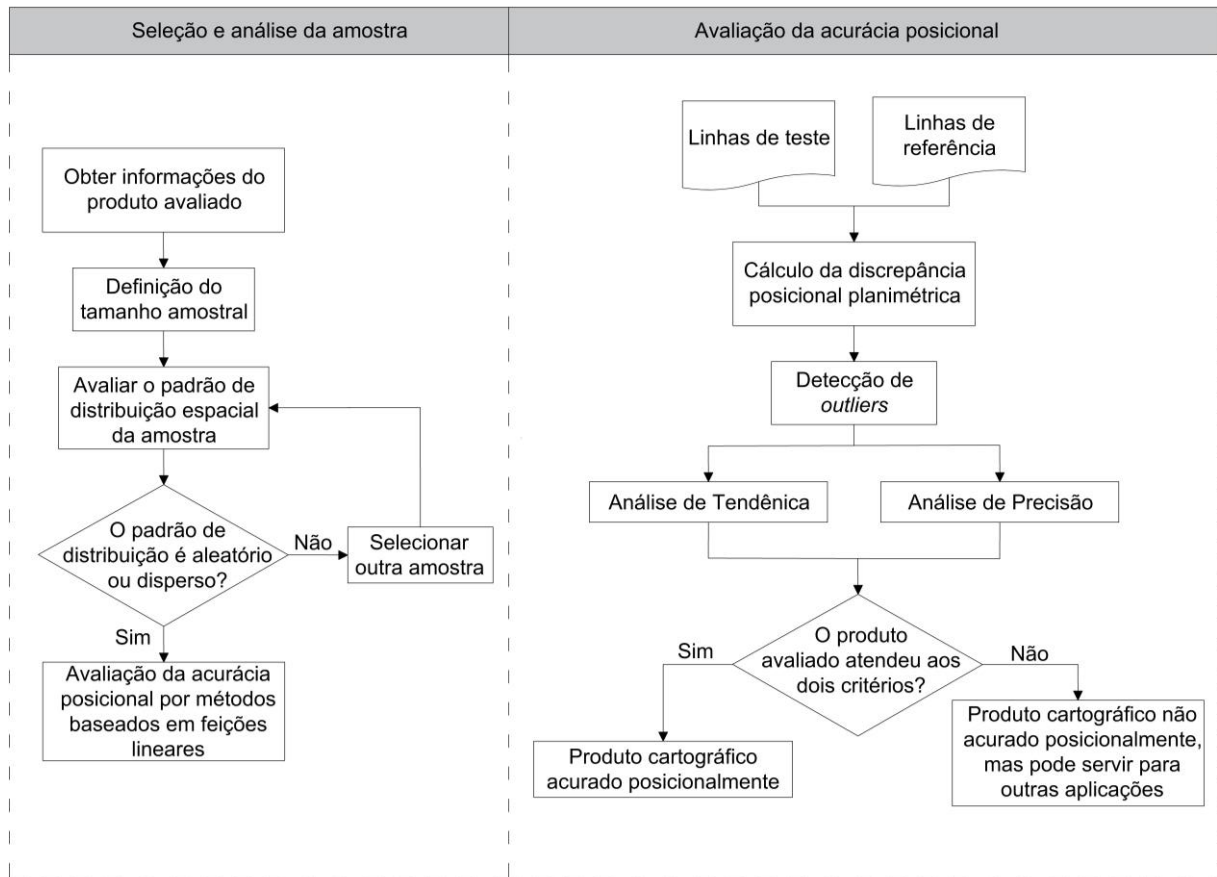


### 3.2 Metodologia Proposta

A organização da metodologia proposta neste estudo foi inspirada no trabalho de Santos et al. (2016), que realiza a avaliação da acurácia posicional por meio da análise da precisão e da tendência utilizando pontos.

Portanto, a primeira etapa da metodologia (Figura 4) envolve a obtenção de informações sobre o produto cartográfico avaliado. Essas informações são relacionadas ao *datum* utilizado, o sistema de projeção, a resolução espacial, técnicas utilizadas na sua produção, entre outras. O conhecimento destas informações fornece uma ideia prévia da escala que o produto espacial poderá ser avaliado e utilizado.

Figura 4 - Metodologia.



Fonte: Adaptado de Santos et al. (2016).

A etapa seguinte da metodologia a ser seguida é a determinação do tamanho da amostra de controle. Para isso, a proposta é que se utilize a Equação 2, obtida no Capítulo 2. Esta equação relaciona o tamanho da área avaliada (**A**) com o tamanho amostral mínimo indicado (**L**).

$$L = 2,0176 * A^{0,5478} \quad (2)$$

Definido o tamanho amostral, o próximo passo consiste na verificação do padrão de distribuição espacial das feições lineares que compõem a amostra. Para isso, deve ser utilizado o Nearest Neighbor Method for Linear Features (NNMLF), apresentado no Capítulo 1. Nesta etapa podem ser utilizados o *plugin* do QGIS ou o código em linguagem R disponibilizados. Segundo Santos (2010), no controle de qualidade posicional é recomendável que a amostra possua uma distribuição espacial que seja aleatória ou dispersa. Portanto, se o padrão de distribuição espacial da

amostra indicado pelo NNMLF for agrupado, deve-se selecionar outra amostra e reavaliada a sua distribuição espacial, até que o padrão recomendado seja atendido.

Após a análise do padrão de distribuição espacial das feições de checagem, pode-se iniciar a avaliação da acurácia posicional propriamente dita, que compreende a segunda etapa da metodologia, que se inicia com o cálculo das discrepâncias planimétricas. Para esse propósito, podem ser utilizados os métodos descritos no Tópico 2 do Capítulo 2: *Buffer* Simples (GOODCHILD e HUNTER, 1997), *Buffer* Duplo (TVEITE e LANGAAS, 1999), Distância de Hausdorff (ABBAS, 1994) e Influência do Vértice (MOZAS-CALVACHE e ARIZA-LÓPEZ, 2011). Estes quatro métodos foram utilizados nesta pesquisa. Santos et al. (2015) apresentam uma metodologia para utilizar as discrepâncias provenientes deste para avaliação segundo a norma brasileira. Portanto, sugere-se seguir estas orientações.

O próximo passo é verificar a presença de *outliers* no conjunto de discrepâncias obtido. Para isso, pode ser utilizado o método  $3\sigma$ . Este método é um dos mais conhecidos e utilizados para identificar candidatos a *outliers* (CRISTO, 2021).

A premissa deste teste é que *outliers* raramente acontecem. Dessa forma, observações fora do intervalo  $\pm 3\sigma$  podem ser consideradas como possíveis *outliers* (GHILANI, 2010). De fato, deve-se tratar essas observações *a priori* como possíveis ou candidatos a *outliers*. Se algum valor de discrepância for tido como *outlier*, deve-se observar se as feições foram identificadas corretamente nos produtos de teste e de referência. Se a identificação apresentar alguma inconsistência (erro grosseiro), deve-se tentar corrigir o erro ou então descartar esse dado. Entretanto, se não houver nenhum problema com este dado, ele deve ser mantido no conjunto de discrepâncias, pois ele reflete a qualidade posicional do produto avaliado. Sua retirada poderia interferir no processo de avaliação (SANTOS et al., 2016).

Partindo dessa premissa, uma discrepância  $d_i$  é uma candidata a *outlier* se, em módulo, seu valor for maior que três vezes o valor do desvio-padrão amostral, conforme Equação 3. Ghilani (2010) alerta que, apesar do nome do método ser ‘ $3\sigma$ ’, este se baseia no desvio-padrão amostral ( $s$ ), e não no desvio-padrão populacional ( $\sigma$ ).

$$|d_i| > 3 * s \quad (3)$$

Para a aplicação deste método no Controle de Qualidade Cartográfica, Nero (2005) recomenda adotar  $3*EP$  como tolerância. Portanto, o critério para detecção de candidatos a *outliers*, utilizado na Cartografia, é dado pela Equação 4. Uma vantagem dessa formulação é que, por utilizar o Erro-Padrão, o critério de detecção está relacionado com a escala de avaliação do produto cartográfico, não sendo um valor estático.

$$|d_i| > 3 * EP \quad (4)$$

A última etapa do processo de avaliação da acurácia posicional envolve as análises de precisão e de tendência. Para a análise de tendência, pode ser utilizado o método TRS (*Translation, Rotation and Scale*), proposto por Zanetti (2022), que avalia a translação, a rotação e a escala de feições lineares.

O método TRS utiliza o método da Média Direcional e Variância Circular para a detecção do efeito de translação. Santos et al. (2016) propuseram a combinação destas estatísticas espaciais para identificar efeitos sistemáticos em pontos. Entretanto, Zanetti (2022) aplica estas estatísticas em feições lineares.

Neste método é realizado o traçado de vetores ligando os vértices homólogos das linhas de referência ( $L_R$ ) e teste ( $L_T$ ). Em seguida, é calculada a Média Direcional ( $\theta_R$ ) (Equação 5), obtida através do azimute ( $\theta_v$ ) dos vetores das discrepâncias posicionais dos vértices.

$$\theta_R = \arctg\left(\frac{\sum \text{sen}(\theta_v)}{\sum \text{cos}(\theta_v)}\right) \quad (5)$$

A Média Direcional apenas determina a tendência da direção dos vetores de discrepâncias. Para se obter a variabilidade desta direção é necessário calcular a Variância Circular ( $S_c$ ) (SANTOS et al., 2016). A Equação 6 fornece a estimativa do comprimento do vetor resultante ( $C_R$ ) (WONG e LEE, 2005; SANTOS, 2010; SANTOS et al., 2016).

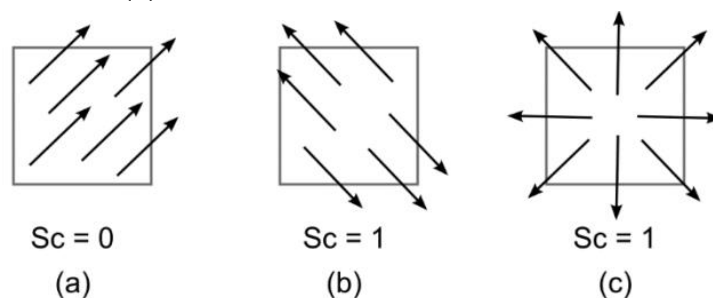
$$C_R = \sqrt{(\sum \text{sen } \theta_v)^2 + (\sum \text{cos } \theta_v)^2} \quad (6)$$

Por fim, a Variância Circular ( $S_C$ ) é obtida por meio do comprimento do vetor resultante ( $C_R$ ) e do número de vetores ( $n$ ), conforme Equação 7.

$$S_C = 1 - \frac{C_R}{n} \quad (7)$$

Segundo Santos et al. (2016), se todos os vetores apresentarem a mesma direção, ou seja, a Variância Circular é nula ( $S_C = 0$ , Figura 5a), a Média Direcional é representativa, indicando a presença de tendência; translação, neste caso. Por outro lado, se estas direções são opostas ( $S_C = 1$ , Figura 5b), a Média Direcional não é representativa e, portanto, não há indícios de efeitos sistemáticos. O caso exemplificado na Figura 5c ilustra uma situação de tendência radial, não detectada pela Média Direcional e Variância Circular. Entretanto, para a análise de translação, não faz sentido utilizar a Média Direcional e a Variância Circular nesta situação. Possivelmente este caso será detectado como efeito de escala, que será descrito a seguir.

Figura 5 – Análise da tendência (translação) utilizando a Média Direcional e a Variância Circular. (a) Vetores na mesma direção, (b) vetores em direções opostas e (c) vetores com tendência radial.



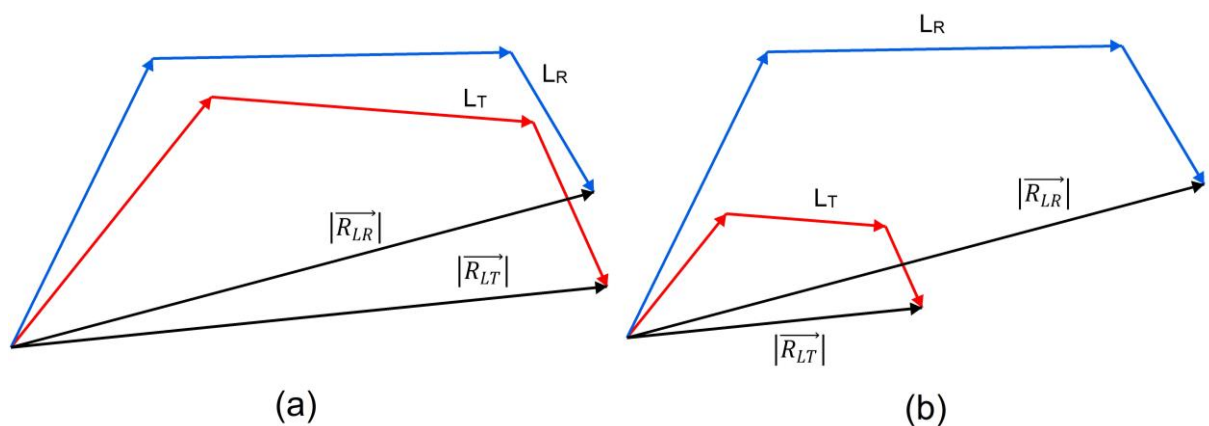
Fonte: Santos et al. (2016).

Para a análise da presença de tendência, pode-se adotar o mesmo critério de Zanetti (2022). Se a Variância Circular for menor ou igual a 0,5 ( $S_C \leq 0,5$ ), pode-se dizer que o produto cartográfico avaliado apresenta tendência de translação. Caso contrário, pode-se afirmar que o produto não apresenta esta tendência.

A segunda etapa na análise de tendência é a verificação da presença do efeito de escala, realizada por meio da Análise Vetorial conforme proposto por Zanetti (2022). A premissa básica é que, se não houver efeito de escala no produto avaliado,

a linha de teste ( $L_T$ ) e a linha de referência ( $L_R$ ) terão a mesma resultante ( $\vec{R}$ ), como representado na Figura 6a. Esta resultante é dada pelo módulo do vetor que liga os vértices inicial e final de cada linha de teste e de referência. No caso da Figura 6b, percebe-se que o módulo do vetor resultante da linha de teste ( $|\vec{R}_{LT}|$ ) é diferente do módulo do vetor resultante da linha de referência ( $|\vec{R}_{LR}|$ ). Nesta figura, a linha de teste ( $L_T$ ) apresenta um efeito de escala em relação a linha de referência ( $L_R$ ).

Figura 6 - Efeito de escala a partir da Análise Vetorial. (a) Sem efeito de escala e (b) com efeito de escala.



Fonte: Adaptado de Zanetti (2022).

Portanto, para análise do efeito de escala, deve ser calculado o Fator de Escala ( $FE$ ), que é dado pela razão entre o módulo do vetor resultante da linha de teste ( $|\vec{R}_{LT}|$ ) e o módulo do vetor resultante da linha de referência ( $|\vec{R}_{LR}|$ ) (Equação 8).

$$FE = \frac{|\vec{R}_{LR}|}{|\vec{R}_{LT}|} \quad (8)$$

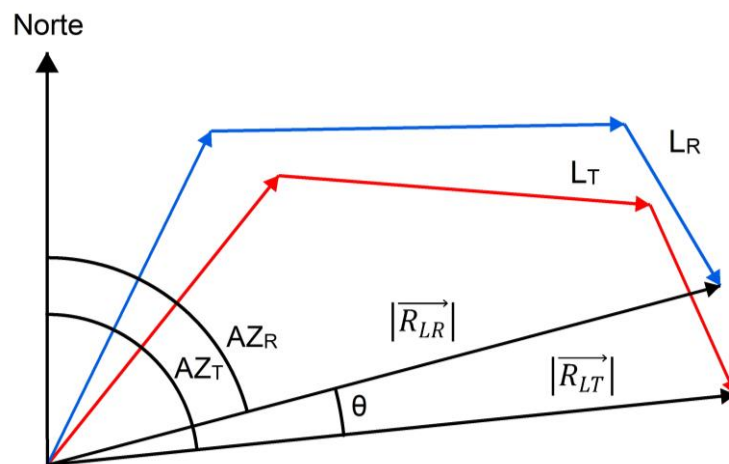
Segundo Zanetti (2022), se o Fator de Escala de todas as linhas avaliadas estiver entre 0,9 e 1,1 ( $0,9 \leq FE \leq 1,1$ ), pode-se afirmar que o produto cartográfico avaliado não apresenta efeito de escala. Entretanto, se alguma linha não satisfizer essa condição novas análises são necessárias. Neste caso, deve-se calcular a média ponderada ( $MP$ ) do conjunto de Fatores de Escala, tendo como pesos os comprimentos das linhas de referência ( $L_R$ ), como apresentado pela Equação 9.

$$MP = \frac{\sum(FE * |\overline{R_{LR}}|)}{\sum|\overline{R_{LR}}|} \quad (9)$$

De forma semelhante, se a média ponderada (**MP**) das linhas estiver entre 0,9 e 1,1 ( $0,9 \leq MP \leq 1,1$ ), pode-se afirmar que o produto cartográfico avaliado não apresenta efeito de escala. Caso contrário, o produto apresenta tendência com efeito de escala.

A última etapa da avaliação de tendência é a análise de rotação, que Zanetti (2022) também propõe uma Análise Vetorial. Para isso, a autora utiliza a Lei dos Cossenos para obter o ângulo  $\theta$  formado entre os vetores resultantes das linhas de teste e de referência. Entretanto, como pode ser observado na Figura 7, este ângulo pode ser determinado por uma relação entre os azimutes dos vetores resultantes (Equação 10). Assim como no caso da utilização da Lei dos Cossenos, é necessário realizar uma análise de quadrantes para se evitar valores de  $\theta$  negativos.

Figura 7 - Ângulo de rotação.



$$\theta = AZ_R - AZ_T \quad (10)$$

Obtido o conjunto de ângulos  $\theta$  para todas as linhas avaliadas, deve-se calcular a média ponderada (**MP**), conforme Equação 11. Neste caso, os pesos são os comprimentos das linhas de referência (**C<sub>LR</sub>**).

$$MP = \frac{\sum(\text{sen}(\theta) * C_{LR})}{\sum C_{LR}} \quad (11)$$

Da mesma forma que ocorre na análise de escala, se a média ponderada (**MP**) das linhas estiver entre 0,9 e 1,1 ( $0,9 \leq MP \leq 1,1$ ), pode-se afirmar que o produto cartográfico avaliado não apresenta efeito de rotação. Caso contrário, o produto apresenta tendência com efeito de rotação.

Por fim, para a análise de precisão, inicialmente se deve estabelecer para qual escala e classe (Tabela 2) o produto será avaliado, para se obter os valores de PEC-PCD e EP. A partir desses valores e do conjunto de discrepâncias posicionais, duas condições devem ser atendidas:

- (i) Noventa por cento (90%) dos valores de discrepâncias posicionais devem ser menores ou iguais ao valor de PEC-PCD, para a classe e a escala testadas; e
- (ii) O RMS (*Root Mean Square*) do conjunto de discrepâncias deve ser menor ou igual ao valor de EP, para a classe e escala testadas.

Se o produto cartográfico avaliado for aprovado nas análises de tendência (não apresentar translação, rotação e escala) e de precisão simultaneamente, pode-se dizer que ele é acurado posicionalmente para a classe e escala testadas. Entretanto, se o produto for reprovado em pelo menos uma das análises, ele ainda pode ser destinado a outra aplicação. Se for detectado algum efeito sistemático, este pode ser corrigido ou minimizado, e o processo deve ser refeito.

#### 4 RESULTADOS E DISCUSSÕES

Visando verificar a eficácia da metodologia proposta, esta foi aplicada a três bases cartográficas: de grande, média e pequena escalas, apresentadas no Tópico 3.1 deste capítulo. Para comparação, estas mesmas bases de dados foram avaliadas utilizando-se o método tradicional por pontos, amplamente utilizado no Brasil e no mundo. No método por pontos foi utilizada a metodologia de Santos et al. (2016). Por serem base de dados de feições lineares, os pontos utilizados foram obtidos a partir das interseções das estradas.

Dessa forma, a base de dados da Catalunha (Figura 1), Espanha, corresponde ao produto de grande escala, as vias do Rio de Janeiro (Figura 2) compõem o produto cartográfico de média escala e as estradas Austrália (Figura 3) representam o produto de pequena escala.

Após a obtenção de informações sobre o produto a ser avaliado, a primeira etapa das duas metodologias (feições lineares e pontuais) envolve a determinação do tamanho amostral. Na avaliação baseada em pontos, segundo DSG (2016a), o produto avaliado deve ser particionado em células de 4 x 4 cm na escala avaliada. Após essa divisão, devem ser analisadas quais células possuem pelo menos um ponto bem identificável; estas células são consideradas válidas. O conjunto de células validadas é denominado de população. A ET-CQDG utiliza as tabelas das normas ISO 2859-1 e ISO 2859-2 para a determinação do tamanho da amostra a partir do tamanho da população.

Na primeira área de estudo, com extensão de 14 km<sup>2</sup> na Catalunha (Figura 1) e escala avaliada de 1/5.000, houveram 361 células válidas. A segunda área, com 43750 km<sup>2</sup> no Rio de Janeiro (Figura 2) e escala avaliada de 1/250.000, contou com 471 células. E a terceira área, no nordeste da Austrália (Figura 3), cujos dados foram avaliados na escala de 1/5.000.000, e extensão de 944530 km<sup>2</sup>, contou com 30 células. Segundo a ET-CQDG, os tamanhos amostrais destas áreas, a partir de suas populações, são de 20, 20 e 10 pontos, respectivamente.

Já a determinação do tamanho amostral na metodologia proposta para avaliação da acurácia posicional utilizando feições lineares é dada pela Equação 2. Esta equação estabelece uma relação direta entre o tamanho da área avaliada e o tamanho amostral mínimo recomendado. Para as áreas de 14 km<sup>2</sup>, 43750 km<sup>2</sup> e 944530 km<sup>2</sup>, os tamanhos amostrais indicados são 8,5 km, 703,3 km e 3784,9 km, respectivamente. Um resumo com estas informações é apresentado na Tabela 3.

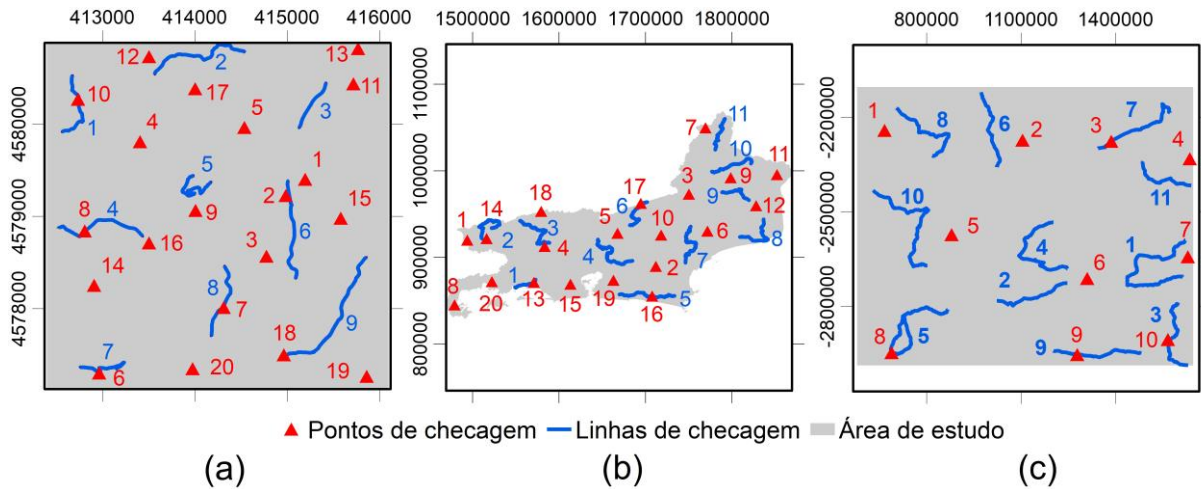
Tabela 3 - Tamanho amostral para a avaliação da acurácia posicional utilizando pontos e linhas.

|                                  | Área (km <sup>2</sup> ) | Escala avaliada | Tamanho amostral |             |
|----------------------------------|-------------------------|-----------------|------------------|-------------|
|                                  |                         |                 | Pontos           | Linhas (km) |
| Catalunha<br>(Grande escala)     | 14                      | 1/5.000         | 20               | 8,5         |
| Rio de Janeiro<br>(Média escala) | 43750                   | 1/250.000       | 20               | 703,3       |
| Austrália<br>(Pequena escala)    | 944530                  | 1/5.000.000     | 10               | 3784,9      |

Observando estes resultados, percebe-se que as regiões da Catalunha e do Rio de Janeiro, apesar de apresentarem extensões de área significativamente diferentes (uma área 3000 vezes maior que a outra), apresentam o mesmo tamanho amostral na metodologia baseada em pontos. Esta pode ser uma desvantagem dessa metodologia. É natural associar um produto cartográfico em grande escala com um mapeamento que abrange uma área pequena. E, da mesma forma, é comum relacionar um mapeamento de grande extensão territorial com uma escala pequena. Entretanto, esta não é uma regra. Com o avanço da tecnologia, os equipamentos utilizados no mapeamento/levantamento têm ganhado um aumento na produtividade, permitindo mapear áreas cada vez maiores em grande escala. Um exemplo é a própria base de dados da Catalunha, cuja escala é de 1/1.000 e possui extensão total de 32368 km<sup>2</sup>. Esta base foi obtida através de um voo fotogramétrico utilizando uma câmara métrica Leica ADS40, sistemas GNSS (*Global Navigation Satellite System*) e inercial, e plataforma giroestabilizada (ICGC, 2012).

As amostras lineares e pontuais obtidas são apresentadas na Figura 8.

Figura 8 - Amostras pontuais e lineares utilizadas. (a) Catalunha, (b) Rio de Janeiro e (c) Austrália.



A próxima etapa das metodologias consistiu em avaliar o padrão de distribuição espacial destas amostras. Para a amostra pontual foi utilizado o tradicional Método do Vizinho mais Próximo. Já para o caso das linhas, foi utilizado o método Nearest Neighbor Method for Linear Features (NNMLF), proposto no Capítulo 1. As Tabelas 4 e 5 apresentam os valores de distância média observada ( $R_{OBS}$ ), distância média esperada ( $R_{ESP}$ ), índice  $R$ , tamanho da área ( $A$ ), valores de  $Z$  calculado e tabelado, e o resultado, para as três áreas de estudo. Neste estudo foi verificada somente a primeira ordem, para os dois métodos.

Tabela 4 - Resultado do Método do Vizinho mais Próximo para pontos.

|                        | Pontos                   |                               |                          |
|------------------------|--------------------------|-------------------------------|--------------------------|
|                        | Catalunha<br>(Figura 8a) | Rio de Janeiro<br>(Figura 8b) | Austrália<br>(Figura 8c) |
| $R_{OBS}$              | 634,853                  | 42053,706                     | 300891,913               |
| $R_{ESP}$              | 417,674                  | 23385,474                     | 153666,065               |
| $R$                    | 1,520                    | 1,798                         | 1,958                    |
| $Z$                    | 4,449                    | 6,830                         | 5,796                    |
| $Z$ tabelado           |                          | 1,960                         |                          |
| $A$ (km <sup>2</sup> ) | 14                       | 43750                         | 944530                   |
| Nível de Confiança     |                          | 95%                           |                          |
| Resultado              | Disperso                 | Disperso                      | Disperso                 |

Tabela 5 - Resultado do Nearest Neighbor Method for Linear Features (NNMLF).

|                           | Linhas                   |                               |                          |
|---------------------------|--------------------------|-------------------------------|--------------------------|
|                           | Catalunha<br>(Figura 8a) | Rio de Janeiro<br>(Figura 8b) | Austrália<br>(Figura 8c) |
| <b>R<sub>OB</sub>S</b>    | 1076,683                 | 45631,591                     | 184366,258               |
| <b>R<sub>ESP</sub></b>    | 623,610                  | 31532,811                     | 146514,815               |
| <b>R</b>                  | 1,727                    | 1,447                         | 1,266                    |
| <b>Z</b>                  | 4,171                    | 2,838                         | 1,686                    |
| <b>Z tabelado</b>         |                          | 1,960                         |                          |
| <b>A (km<sup>2</sup>)</b> | 14                       | 43750                         | 944530                   |
| Nível de Confiança        |                          | 95%                           |                          |
| Resultado                 | Disperso                 | Disperso                      | Aleatório                |

Como pode ser observado nestas tabelas, todas as amostras apresentaram padrão de distribuição espacial ou disperso, que é situação recomendável de uma amostra utilizada no Controle de Qualidade Cartográfica. Caso uma amostra apresentasse o padrão agrupado, outros elementos amostrais deveriam ser escolhidos e o processo deveria ser refeito. Como isso não ocorreu, pôde-se obter as discrepâncias posicionais das amostras.

No caso das linhas, essas discrepâncias foram obtidas utilizando-se os métodos *Buffer Simples* (BS), *Buffer Duplo* (BD), Distância de Hausdorff (DH) e Influência do Vértice (IV), seguindo as metodologias descritas no Capítulo 2. Esses métodos foram escolhidos por serem os mais utilizados na avaliação da acurácia posicional utilizando feições lineares. Já para os pontos, foi calculada a discrepância posicional planimétrica ( $d_{2D}$ ) entre as coordenadas dos pontos de referência ( $X_T, Y_T$ ) e de teste ( $X_R, Y_R$ ), conforme a Equação 12.

$$d_{2D} = \sqrt{(X_T - X_R)^2 + (Y_T - Y_R)^2} \quad (12)$$

Os valores das discrepâncias obtidas a partir das amostras da base de dados da Catalunha (Figura 8a), avaliada na escala de 1/5.000, são apresentadas na Figura 9 (amostras pontuais) e na Figura 10 (amostras lineares). Nestas figuras, o eixo X corresponde ao identificador único (ID) de cada feição, mostrado na Figura 8. As estatísticas descritivas destes conjuntos de discrepâncias encontram-se na Tabela 6.

Figura 9 - Discrepâncias obtidas pelo método baseado em pontos (Catalunha).

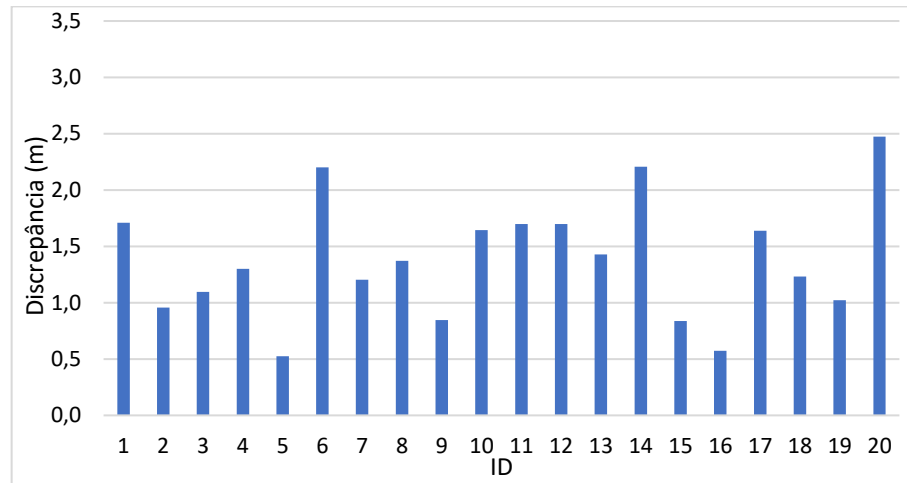


Figura 10 - Discrepâncias obtidas pelo método baseado em linhas (Catalunha).

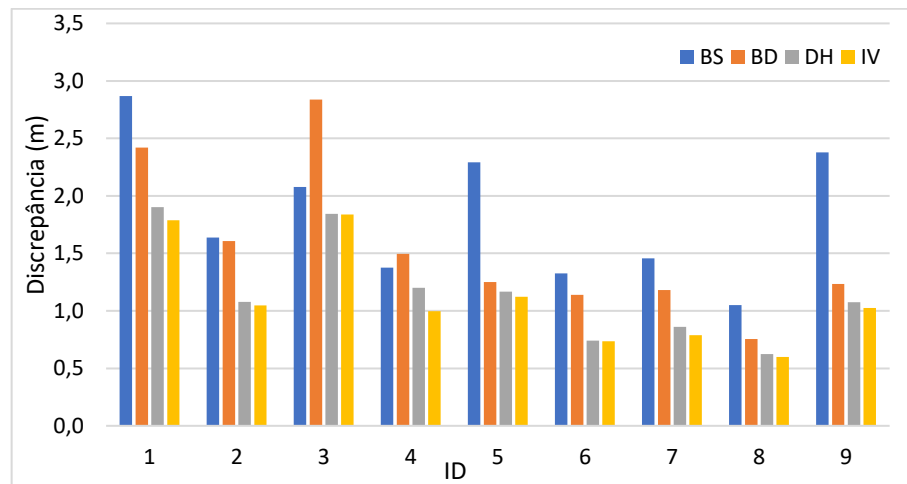


Tabela 6 – Estatísticas descritivas das discrepâncias da área da Catalunha.

| Estatística (m) | Pontos | Linhas |       |       |       |
|-----------------|--------|--------|-------|-------|-------|
|                 |        | BS     | BD    | DH    | IV    |
| Mínimo          | 0,525  | 1,050  | 0,755 | 0,625 | 0,598 |
| Máximo          | 2,474  | 2,869  | 2,837 | 1,901 | 1,838 |
| Média           | 1,383  | 1,829  | 1,546 | 1,166 | 1,104 |
| Desvio-padrão   | 0,534  | 0,602  | 0,666 | 0,445 | 0,436 |
| RMS             | 1,478  | 2,031  | 1,770 | 1,315 | 1,250 |

Já para a área de estudo do Rio de Janeiro, cuja base de dados foi avaliada na escala de 1/250.000, as Figuras 11 e 12 apresentam as discrepâncias obtidas pelos métodos pontuais e lineares, respectivamente. A Tabela 7 contém as estatísticas desses resultados.

Figura 11 - Discrepâncias obtidas pelo método baseado em pontos (Rio de Janeiro).

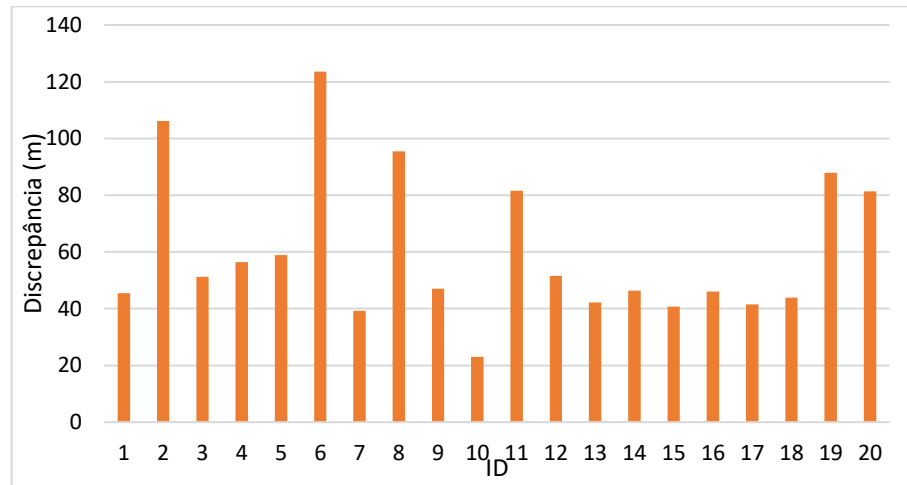


Figura 12 - Discrepâncias obtidas pelo método baseado em linhas (Rio de Janeiro).

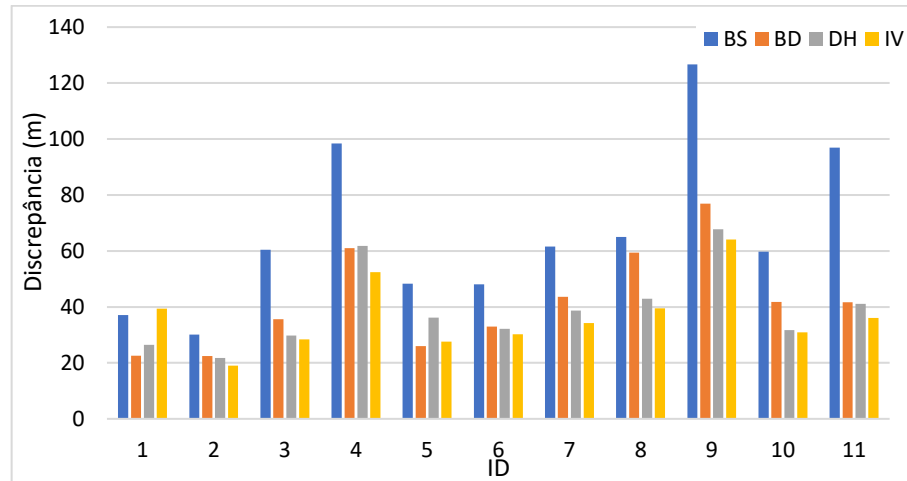


Tabela 7 – Estatísticas descritivas das discrepâncias da área do Rio de Janeiro.

| Estatística (m) | Pontos  | Linhas  |        |        |        |
|-----------------|---------|---------|--------|--------|--------|
|                 |         | BS      | BD     | DH     | IV     |
| Mínimo          | 22,954  | 30,158  | 22,503 | 21,740 | 18,971 |
| Máximo          | 123,600 | 126,689 | 76,873 | 67,802 | 64,119 |
| Média           | 60,485  | 66,585  | 42,173 | 39,129 | 36,536 |
| Desvio-padrão   | 26,304  | 29,215  | 17,368 | 14,210 | 12,499 |
| RMS             | 65,694  | 75,699  | 47,520 | 43,429 | 40,306 |

Por fim, para a área de estudo da Austrália, cuja base de dados foi avaliada segundo a escala de 1/5.000.000, as discrepâncias obtidas podem ser observadas na Figura 13 (pontos) e na Figura 14 (linhas). As estatísticas destas discrepâncias são apresentadas na Tabela 8.

Figura 13 - Discrepâncias obtidas pelo método baseado em pontos (Austrália).

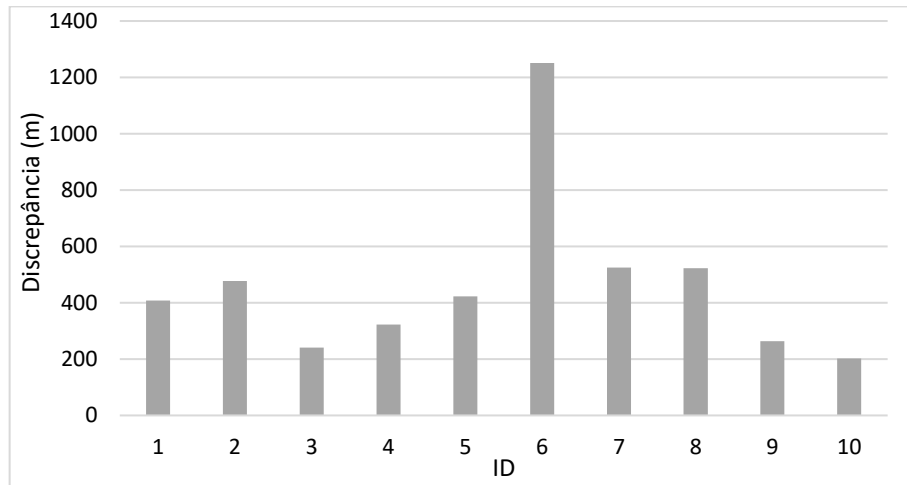


Figura 14 - Discrepâncias obtidas pelo método baseado em linhas (Austrália).

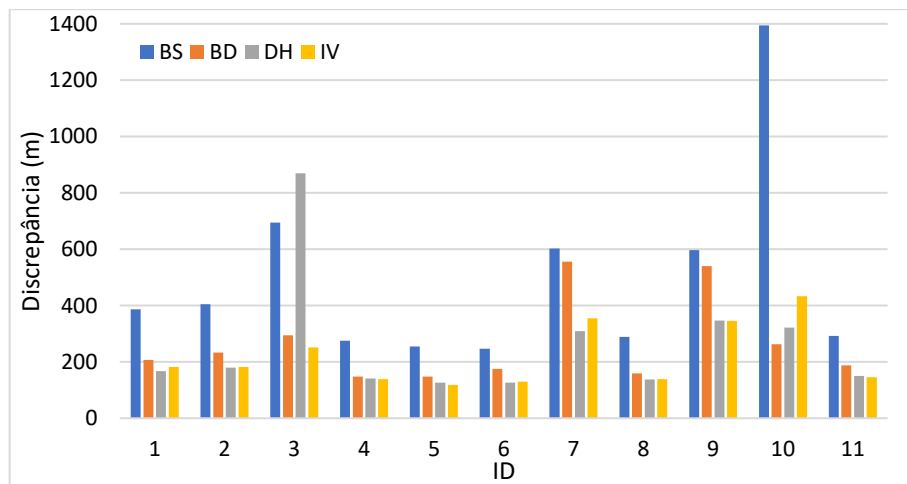
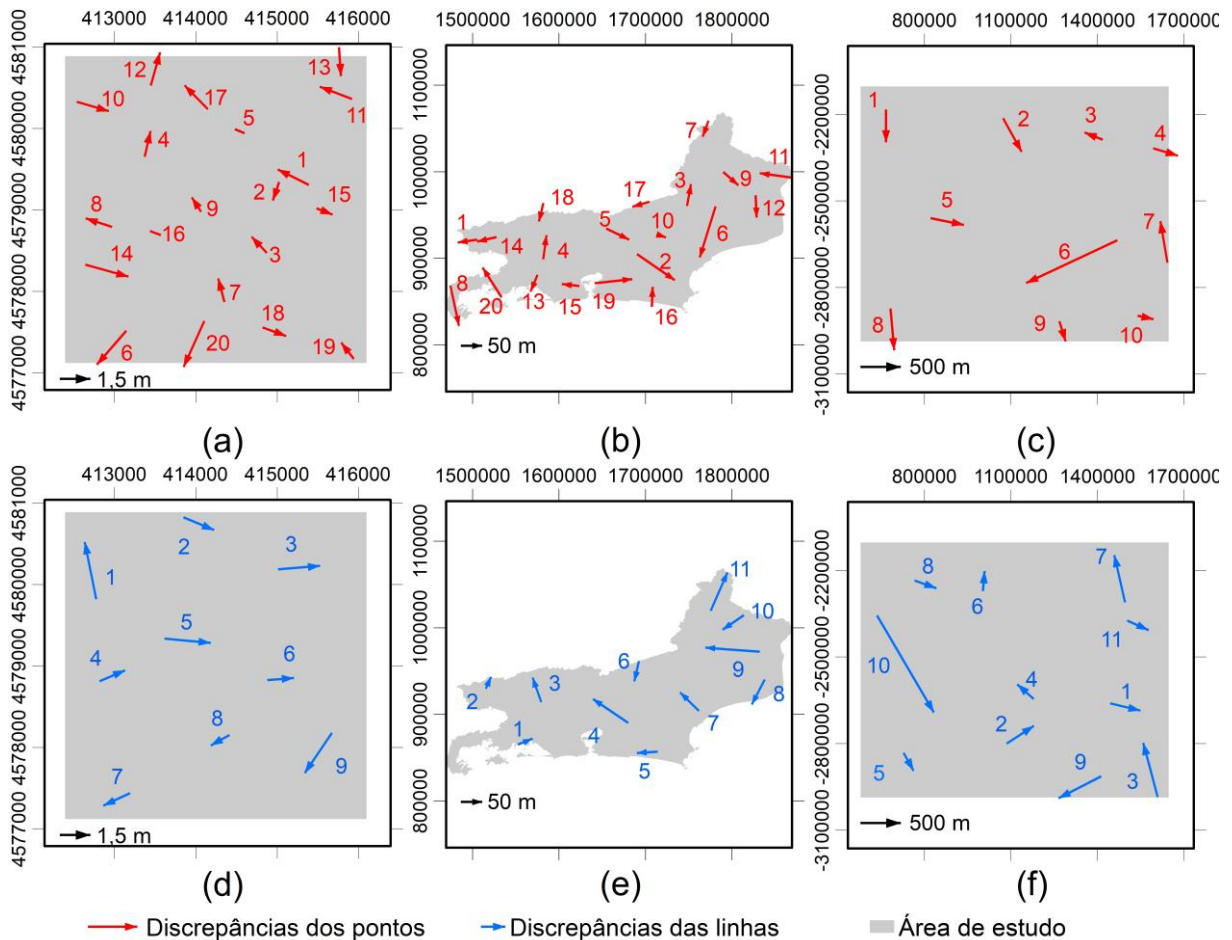


Tabela 8 – Estatísticas descritivas das discrepâncias da área da Austrália.

| Estatística (m) | Pontos   | Linhas   |         |         |         |
|-----------------|----------|----------|---------|---------|---------|
|                 |          | BS       | BD      | DH      | IV      |
| Mínimo          | 202,104  | 247,263  | 148,053 | 126,461 | 117,902 |
| Máximo          | 1251,273 | 1394,458 | 555,803 | 869,147 | 432,575 |
| Média           | 463,352  | 494,315  | 264,822 | 261,332 | 219,860 |
| Desvio-padrão   | 300,104  | 337,849  | 147,622 | 218,375 | 109,653 |
| RMS             | 543,830  | 618,809  | 314,541 | 350,444 | 255,336 |

Para as três bases de dados analisadas, em diferentes escalas, o método do *Buffer Simples* (BD) apresentou as maiores discrepâncias (tanto em média quanto em valores absolutos) e o maior RMS quando comparado com os demais métodos de feição linear e o método por pontos. Em seguida, o método que forneceu as maiores discrepâncias foi o método baseado em pontos; apresentando discrepâncias médias e máximas superior aos demais métodos em quase todas as análises. Este resultado é semelhante ao obtido por Cunha et al. (2019), que compararam os métodos *Buffer Simples*, *Buffer Duplo* e *Influência do Vértice*. A Figura 15 traz a espacialização dos vetores de discrepância posicional obtidos com os métodos por ponto e por linhas. Pelo método do *Buffer Simples* ter apresentado as maiores discrepâncias, somente o seu resultado é apresentado.

Figura 15 – Vetores de discrepância posicional. (a) Grande escala – pontos, (b) Média escala – pontos, (c) Pequena escala – pontos, (d) Grande escala – linhas, (e) Média escala – linhas, e (f) pequena escala - linhas.



Mas, não é uma tarefa simples comparar as discrepâncias obtidas por métodos baseado em tipos de feições diferentes. A própria maneira do *Buffer* Simples obter as discrepâncias parece conduzir a maiores valores. A discrepância posicional deste método é dada pela largura do *buffer* da linha de referência que ocasiona 90% de inclusão da linha de teste. Então, enquanto esta porcentagem não foi atingida, a discrepância posicional irá aumentar. Por outro lado, os demais métodos de feição linear apresentam algumas características que podem atenuar os valores das discrepâncias.

O *Buffer* Duplo utiliza uma relação de áreas contidas ou não nos *buffers* das linhas de teste e de referência. Devido ao fato de o *buffer* ser aplicado em ambas as linhas (referência e teste), a condição exigida para se interromper o processo iterativo é obtida mais rapidamente, o que pode conduzir a um valor de largura de *buffer* ideal menor, amenizando as discrepâncias obtidas. Por fim, os métodos da Distância de Hausdorff e da Influência do Vértice são baseados nas menores distâncias entre as linhas de referência e de teste; atenuando os valores de discrepâncias destes métodos. Portanto, nenhum dos métodos de feição linear utilizados, nem mesmo o *Buffer* Simples, apresenta os valores máximos de afastamento entre as linhas.

Já o método baseado em pontos, por realizar uma análise única, pontual, depende da localização da amostra. Essa característica do método pode levar a valores de discrepâncias críticos (tanto menores quanto maiores) se comparado com os métodos de feição linear. Considere a Figura 16, que ilustra um caso real ocorrido na base de dados da Catalunha.



m, respectivamente. Já os maiores valores de discrepância, como pode ser observado nas Tabelas 6, 7 e 8, foram 2,869 m, 126,689 m e 1394,458 m, respectivamente.

Diante disso, foi possível seguir com as análises de tendência e precisão. Para a análise de tendência na metodologia baseada em pontos, Santos et al. (2016) propõe utilizar a Média Direcional e a Variância Circular, semelhante a proposta de Zanetti (2022) para a detecção do efeito sistemático de translação em feições lineares. No caso dos pontos, os valores de Variância Circular (**Sc**) obtidos são apresentados na Tabela 9.

Tabela 9 - Análise de tendência para as amostras pontuais.

|                        | Catalunha<br>(1/5.000) | Rio de Janeiro<br>(1/250.000) | Austrália<br>(1/5.000.000) |
|------------------------|------------------------|-------------------------------|----------------------------|
| <b>Sc</b> (translação) | 0,9166                 | 0,7712                        | 0,5995                     |

Como nenhum valor de **Sc** foi inferior a 0,5, este teste apontou que as três bases de dados não apresentam tendência.

Para a análise da tendência nas amostras lineares, o método TRS (*Translation, Rotation and Scale*) foi implementado na linguagem R. Este código encontra-se disponível no Apêndice A. A Tabela 10 apresenta os principais valores para a análise de translação, rotação e escala

Tabela 10 - Análise de tendência para as amostras lineares.

|                           | Catalunha<br>(1/5.000) | Rio de Janeiro<br>(1/250.000) | Austrália<br>(1/5.000.000) |
|---------------------------|------------------------|-------------------------------|----------------------------|
| <b>Sc</b> (translação)    | 0,5569                 | 0,8620                        | 0,7912                     |
| <b>FE</b> mínimo (escala) | 0,9806                 | 0,9717                        | 0,9843                     |
| <b>FE</b> máximo (escala) | 1,0559                 | 1,0199                        | 1,0011                     |
| <b>MP</b> (rotação)       | 0,0732                 | 0,0068                        | 0,0077                     |

Assim como no caso dos pontos, todos os valores de Variância Circular foram superiores a 0,5, indicando que as amostras das três áreas de estudo utilizadas não apresentam efeito de translação.

Para o caso da escala, todos os valores de Fator de Escala (**FE**) se mantiveram entre 0,9 e 1,1, como propõe Zanetti (2022), o que pode ser observado pelos valores máximos e mínimos do **FE** presentes na Tabela 10. Devido ao fato de nenhum valor ser superior a 1,1 ou inferior a 0,9, não se fez necessário calcular a média ponderada

do Fator de Escala, e pode-se afirmar que todos os dados utilizados nesta pesquisa não apresentam efeito de escala.

Por fim, a análise de rotação consiste no estudo da média ponderada (**MP**) do ângulo  $\theta$ , formado entre os vetores resultantes das linhas de teste e de referência. Neste caso, todos os valores de **MP** se mantiveram entre -0,1 e 0,1, que é o requisito para se afirmar que os dados não apresentam efeito de rotação.

No geral, as duas metodologias apresentam o mesmo resultado (dados sem tendência). Entretanto, é importante destacar uma das vantagens da metodologia proposta, que a análise dos efeitos de translação, rotação e escala. Apesar de ser possível realizar esta análise utilizando pontos, a metodologia mais empregada para a verificação da qualidade posicional no Brasil não propõe a investigação de efeitos de rotação e escala, o que pode tornar essa menos rigorosa.

A última etapa da metodologia proposta consiste na análise da precisão. Para tal, os conjuntos de discrepâncias e os valores de RMS obtidos anteriormente foram comparados com os valores de PEC-PCD e EP da ET-CQDG. Os resultados desta etapa estão dispostos nas Tabelas 11, 12 e 13. Nestas tabelas são indicados dois parâmetros:

(1) Porcentagem de discrepâncias menores ou iguais ao valor do PEC-PCD, para a classe e a escala avaliadas.

(2) Verificação se o RMS do conjunto de discrepâncias é menor que o valor de EP, para a classe e a escala avaliadas.

Tabela 11 - Análise da precisão para os dados da Catalunha. Escala de avaliação: 1/5.000.

| Classe<br>ET-CQDG | Pontos        |               | Linhas        |               |               |               |               |               |               |               |
|-------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
|                   |               |               | BS            |               | BD            |               | DH            |               | IV            |               |
|                   | (1)           | (2)           | (1)           | (2)           | (1)           | (2)           | (1)           | (2)           | (1)           | (2)           |
| A                 | 55,00         | Falhou        | 33,33         | Falhou        | 55,56         | Falhou        | 77,78         | Falhou        | 77,78         | Falhou        |
| B                 | <b>100,00</b> | <b>Passou</b> | 88,89         | Falhou        | 88,89         | Falhou        | <b>100,00</b> | <b>Passou</b> | <b>100,00</b> | <b>Passou</b> |
| C                 | 100,00        | Passou        | <b>100,00</b> | <b>Passou</b> | <b>100,00</b> | <b>Passou</b> | 100,00        | Passou        | 100,00        | Passou        |
| D                 | 100,00        | Passou        | 100,00        | Passou        | 100,00        | Passou        | 100,00        | Passou        | 100,00        | Passou        |

Tabela 12 - Análise da precisão para os dados do Rio de Janeiro. Escala de avaliação: 1/250.000.

| Classe<br>ET-CQDG | Pontos        |               | Linhas        |               |               |               |               |               |               |               |
|-------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
|                   |               |               | BS            |               | BD            |               | DH            |               | IV            |               |
|                   | (1)           | (2)           | (1)           | (2)           | (1)           | (2)           | (1)           | (2)           | (1)           | (2)           |
| A                 | 70,00         | Falhou        | 72,73         | Falhou        | 90,91         | Falhou        | 100,00        | Falhou        | <b>100,00</b> | <b>Passou</b> |
| B                 | <b>100,00</b> | <b>Passou</b> | 90,91         | Falhou        | <b>100,00</b> | <b>Passou</b> | <b>100,00</b> | <b>Passou</b> | 100,00        | Passou        |
| C                 | 100,00        | Passou        | <b>100,00</b> | <b>Passou</b> | 100,00        | Passou        | 100,00        | Passou        | 100,00        | Passou        |
| D                 | 100,00        | Passou        | 100,00        | Passou        | 100,00        | Passou        | 100,00        | Passou        | 100,00        | Passou        |

Tabela 13 - Análise da precisão para os dados da Austrália. Escala de avaliação: 1/5.000.000.

| Classe<br>ET-CQDG | Pontos        |               | Linhas        |               |               |               |               |               |               |               |
|-------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
|                   |               |               | BS            |               | BD            |               | DH            |               | IV            |               |
|                   | (1)           | (2)           | (1)           | (2)           | (1)           | (2)           | (1)           | (2)           | (1)           | (2)           |
| A                 | <b>100,00</b> | <b>Passou</b> | <b>100,00</b> | <b>Passou</b> | <b>100,00</b> | <b>Passou</b> | <b>100,00</b> | <b>Passou</b> | <b>100,00</b> | <b>Passou</b> |
| B                 | 100,00        | Passou        | 100,00        | Passou        | 100,00        | Passou        | 100,00        | Passou        | 100,00        | Passou        |
| C                 | 100,00        | Passou        | 100,00        | Passou        | 100,00        | Passou        | 100,00        | Passou        | 100,00        | Passou        |
| D                 | 100,00        | Passou        | 100,00        | Passou        | 100,00        | Passou        | 100,00        | Passou        | 100,00        | Passou        |

(1) Porcentagem de discrepâncias menores ou iguais ao valor do PEC-PCD, para a classe e a escala avaliadas.

(2) Verificação se o RMS do conjunto de discrepâncias é menor que o valor de EP, para a classe e a escala avaliadas.

Vale ressaltar que, segundo a norma brasileira (Decreto nº 89.817 e ET-CQDG), para um produto ser considerado preciso noventa por cento (90%) dos valores de discrepâncias posicionais devem ser menores ou iguais ao valor de PEC-PCD e o RMS das discrepâncias deve ser menor ou igual ao valor de EP, considerando a classe e a escala avaliadas em ambos os casos.

Diante destas condições, os resultados apresentados nas tabelas anteriores levam às classificações descritas na Tabela 14.

Tabela 14 - Classificação dos produtos segundo o Decreto nº 89.817 e a ET-CQDG.

|        |    | Catalunha<br>(1/5.000) | Rio de Janeiro<br>(1/250.000) | Austrália<br>(1/5.000.000) |
|--------|----|------------------------|-------------------------------|----------------------------|
| Pontos |    | B                      | B                             | A                          |
|        | BS | C                      | C                             | A                          |
| Linhas | BD | C                      | B                             | A                          |
|        | DH | B                      | B                             | A                          |
|        | IV | B                      | A                             | A                          |

Considerando as 12 análises baseadas em feições lineares (três escalas diferentes e quatro métodos), em 11 delas a classificação resultante da metodologia proposta foi igual ou mais rigorosa que o método tradicional, baseado em pontos. Sendo que em 8 análises a metodologia proposta resultou na mesma classificação que o método por pontos, em outras 3 situações a classificação foi mais rigorosa e somente em uma análise o resultado foi menos restritivo que a referência (pontos). Essa exceção ocorreu com o método Influência do Vértice na avaliação da acurácia posicional da base de dados do Rio de Janeiro (escala de 1/250.000).

Estes resultados demonstram, independentemente da escala, a viabilidade de utilização da metodologia proposta para a avaliação da acurácia posicional de produtos cartográficos utilizando feições lineares, sendo uma boa alternativa ao método baseado em pontos.

## 5 CONCLUSÃO

Atualmente, a metodologia baseada em pontos é amplamente utilizada para a avaliação da acurácia posicional. Entretanto, utilizar feições lineares na avaliação da

acurácia posicional apresenta algumas vantagens, como: aproximadamente 80% das feições presentes em bases cartográficas são linhas, geralmente estas feições possuem boa distribuição espacial, e as linhas possuem mais informações geométricas e posicionais que os pontos. Diante disso, nos últimos anos foram desenvolvidos diversos métodos baseados em feições lineares para se obter a discrepância posicional. Porém, além de serem métodos pouco difundidos e utilizados no Brasil, não há padronização, uma metodologia única, para se realizar o controle posicional utilizando linhas como feições de checagem. Aspectos importantes em um processo de amostragem, como o tamanho da amostra e o seu padrão de distribuição espacial sequer eram abordados.

Diante disso, este estudo propôs uma metodologia completa para a avaliação da acurácia posicional utilizando feições lineares, sendo embasada na norma brasileira. Esta metodologia conta com definição do tamanho amostral, avaliação do padrão de distribuição espacial da amostra, detecção de *outliers*, e análises de tendência e de precisão. Dessa forma, todos os aspectos abordados no método por pontos também possuem um equivalente na avaliação por linhas. Além disso, a metodologia proposta, por utilizar o Método TRS, permite a análise de efeitos sistemáticos de escala e de rotação, que não é realizada pelo método tradicional por pontos.

A metodologia proposta foi aplicada a três bases de dados: Catalunha, Rio de Janeiro e Austrália, sendo avaliadas nas escalas de 1/5.000, 1/250.000 e 1/5.000.000 e comparada com o método por pontos. Por utilizar métodos estatisticamente embasados, e pelo fato de as linhas apresentarem mais informações geométricas que os pontos, a metodologia proposta apresentou resultado compatível, ou mais rigoroso, que o método tradicional que utiliza feições pontuais, para as escalas pequena, média e grande.

Com a proposta de uma metodologia bem definida, espera-se que os métodos baseados em linhas possam ser mais difundidos e utilizados. E, além disso, apesar de ser baseada na norma brasileira, esta metodologia pode ser adaptada para outros países, sendo necessário somente adaptar o critério de aceitação ou rejeição para atender às normas locais.

Para trabalhos futuros, pode-se pensar em criar métodos e metodologias para a avaliação da acurácia posicional de dados tridimensionais. Atualmente,

praticamente todos os dados levantados em campo estão apoiados em sistemas GNSS, sendo, portanto, tridimensionais. Dessa forma, faz sentido que a avaliação também seja tridimensional, ao invés de realizar a separação dos dados em componentes planimétrica e altimétrica para avaliação. Um exemplo de dados nesta situação são aqueles da Catalunha, utilizados nesta pesquisa. Estes dados, em grande escala, são tridimensionais e poderiam ser utilizados em experimentos neste sentido. Além disso, outra pesquisa que pode ser desenvolvida é definição do tamanho amostral relacionado com os riscos do usuário e do produtor.

## REFERÊNCIAS BIBLIOGRÁFICAS

ABBAS, I. **Base de Données Vectorielles et Erreur Cartographique. Problèmes Posés Par Le Controle Ponctuel. Une Méthode Alternative Fondée Sur La Distance de Hausdorff: Le Controle Linéaire.** Thèse de Doctorat en Informatique. Université Denis–Diderot, Paris, France, 1994.

ARIZA-LÓPEZ, Francisco Javier; ATKINSON-GORDO, Alan David. Analysis of some positional accuracy assessment methodologies. **Journal of surveying Engineering**, v. 134, n. 2, p. 45-54, 2008.

ARIZA-LÓPEZ, F. J.; GARCÍA-BALBOA, J. L.; RODRÍGUEZ-AVI, J.; ROBLEDO, J. **Guide for the positional accuracy assessment of geospatial data.** Pan American Institute of Geography and History, Occasional Publication # 563, 2021.

BRASIL. Decreto N° 89.817 de 20 de junho de 1984. **Estabelece as Instruções Reguladoras das Normas Técnicas da Cartografia Nacional.** Brasil. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/decreto/1980-1989/D89817.htm](http://www.planalto.gov.br/ccivil_03/decreto/1980-1989/D89817.htm)>. Acesso em: 01 de mar. de 2023. 1984.

CUNHA, M. M.; SECATTO, G. Z.; GALINDO, J. R. F.; SANTOS, A. P. Proposta de um método de avaliação da acurácia posicional baseado na modificação do Buffer Simples. **Revista Brasileira de Cartografia**, v. 71, n. 4, p. 1193-1218, 2019.

CRISTO, Sabrina Lourdes Pereira de. **Análise da detecção e da influência de outliers na avaliação da acurácia posicional de produtos cartográficos.** Dissertação (Mestrado em Estatística Aplicada e Biometria). Viçosa: Universidade Federal de Viçosa. 83p. 2021.

DSG - DIRETORIA DO SERVIÇO GEOGRÁFICO. **Especificação Técnica Para Controle de Qualidade de Dados Geoespaciais (ET-CQDG)**. Brasília, 2016a, 2016.

DSG - DIRETORIA DO SERVIÇO GEOGRÁFICO. **Especificação Técnica Para Produtos de Conjuntos de Dados Geoespaciais (ET-PCDG)**. Brasília, 2016b, 2016.

DSG - DIRETORIA DO SERVIÇO GEOGRÁFICO. **Especificação Técnica Para a Aquisição de Dados Geoespaciais Vetoriais (ET-ADGV)**. Brasília, 2011.

GEOSCIENCE AUSTRALIA. **Global Map**. Austrália: Geoscience Australia, 2001. Shapefile. Escala 1:1.000.000. Disponível em: <https://ecat.ga.gov.au/geonetwork/srv/eng/catalog.search>. Acesso em: 1 dez. 2023.

GEOSCIENCE AUSTRALIA. **GEODATA TOPO**. Austrália: Geoscience Australia, 2004. Shapefile. Escala 1:5.000.000. Disponível em: <https://ecat.ga.gov.au/geonetwork/srv/eng/catalog.search>. Acesso em: 1 dez. 2023.

GOODCHILD, M. F.; HUNTER, G. J. A simple positional accuracy measure for linear features, **International Journal of Geographical Information Science**, 11:3, 299-306, 1997, DOI: 10.1080/136588197242419.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Base Cartográfica Continua**. Rio de Janeiro, Brasil: IBGE, 2023a. Shapefile. Escala 1:25.000. Disponível em: [https://geoftp.ibge.gov.br/cartas\\_e\\_mapas/bases\\_cartograficas\\_continuas/](https://geoftp.ibge.gov.br/cartas_e_mapas/bases_cartograficas_continuas/). Acesso em: 1 dez. 2023.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Base Cartográfica Continua**. Rio de Janeiro, Brasil: IBGE, 2023b. Shapefile. Escala 1:250.000. Disponível em: [https://geoftp.ibge.gov.br/cartas\\_e\\_mapas/bases\\_cartograficas\\_continuas/](https://geoftp.ibge.gov.br/cartas_e_mapas/bases_cartograficas_continuas/). Acesso em: 1 dez. 2023.

ICGC - INSTITUT CARTOGRÀFIC I GEOLÒGIC DE CATALUNYA. **Plec D'especificacions Tècniques: per a l'elaboració de cartografia topogràfica 3D a escales 1:1000 i 1:2000 (CT-1M i CT-2M) v2.2**. Disponível em: <https://datacloud.ide.cat/especificacions/topografia-1000-v2r2-esp-03ca-20120401.pdf>. Acesso: 01 de mar. de 2023. Catalunya, Espanya, 2012.

INSTITUTO CARTOGRÁFICO E GEOLÓGICO DA CATALUNHA. **Cartografia Topográfica 1:1.000**. Catalonia, Espanha: ICGC, 2021a. Shapefile. Escala 1:1.000. Disponível em: <https://www.icgc.cat/>. Acesso em: 1 dez. 2023.

INSTITUTO CARTOGRÁFICO E GEOLÓGICO DA CATALUNHA. **Base Topográfica 1:5.000**. Catalonia, Espanha: ICGC, 2021b. Shapefile. Escala 1:5.000. Disponível em: <https://www.icgc.cat/>. Acesso em: 1 dez. 2023.

ISO - INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 9000:2005: Quality management systems — Fundamentals and vocabulary**. Geneva, Switzerland, 2005.

ISO - INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 19157: Geographic Information—Data Quality**. Geneva, Switzerland, 170 p, 2013.

MOZAS-CALVACHE, A. T.; ARIZA-LÓPEZ, F. J. Methodology for positional quality control in cartography using linear features. **The Cartographic Journal**, v. 47, n. 4, p. 371-378, 2010.

MOZAS-CALVACHE, A. T.; ARIZA-LÓPEZ, F. J. New method for positional quality control in cartography based on lines. A comparative study of methodologies. **International Journal of Geographical Information Science**, v. 25, n. 10, p. 1681-1695. 2011. DOI:10.1080/13658816.2010.545063.

NERO, Marcelo Antônio. **Propostas para o controle de qualidade de bases cartográficas com ênfase na componente posicional**. Tese (Doutorado em Engenharia). Universidade de São Paulo. 2005.

SANTOS, A. P. **Avaliação da Acurácia Posicional em Dados Espaciais com Uso da Estatística Espacial**. Tese (Mestrado em Engenharia Civil) – Universidade Federal de Viçosa, Viçosa, 128p, 2010.

SANTOS, A. P., MEDEIROS, N. G., SANTOS, G. R., RODRIGUES, D. D. Controle de qualidade posicional em dados espaciais utilizando feições lineares. **Boletim de Ciências Geodésicas**, v. 21, n. 2, p. 233-250, 2015.

SANTOS, A. P.; RODRIGUES, D. D.; SANTOS, N. T.; GRIPP JUNIOR, J. Avaliação Da Acurácia Posicional Em Dados Espaciais Utilizando Técnicas De Estatística Espacial: Proposta De Método E Exemplo Utilizando A Norma Brasileira. **Boletim de Ciências Geodésicas**, v. 22, n. 4, p. 630-650, 2016.

SHI, W.; FISHER, P. F.; GOODCHILD, M. F. **Spatial Data Quality**. London and New York: Taylor & Francis, 2005.

TVEITE, Havard; LANGAAS, Sindre. An accuracy assessment method for geographical line data sets based on buffering, **International Journal of Geographical Information Science**, v. 13, n. 1, p. 27-47, 1999, DOI: 10.1080/136588199241445.

WONG, D. W. S; LEE, J. **Statistical analysis of geographic information with ArcView GIS and ArcGIS**. John Wiley & Sons, 2005.

ZANETTI, J. **Método TRS: Uma proposta para avaliar a tendência em produtos cartográficos por meio da análise em feições lineares**. Tese (Doutorado em Engenharia Civil) – Universidade Federal de Viçosa, Viçosa, 119p, 2022.

## CONCLUSÕES GERAIS

Esta pesquisa apresenta métodos para a avaliação do padrão de distribuição espacial e para a determinação do tamanho de amostras utilizadas no controle posicional realizado com feições lineares, o que representa uma inovação na área. Estes temas têm se representado um entrave para a utilização de feições lineares nos processos de controle da qualidade cartográfica, pois impediam a criação de uma metodologia padronizada.

Com a criação destes métodos e reunindo outros já disponíveis na literatura, foi possível elaborar uma metodologia completa para a avaliação da acurácia posicional de feições lineares, para ser utilizada conforme a norma brasileira.

Além de apresentar avanços em questões importantes, este trabalho produziu, utilizou e disponibilizou alguns códigos em linguagem R, muitos deles implementados utilizando programação paralela. São disponibilizados os seguintes códigos: Nearest Neighbor Method for Linear Features (NNMLF), *Buffer* Simples, *Buffer* Duplo, Distância de Hausdorff, Influência do Vértice, Método de seleção aleatória de amostra, Simulação de Monte Carlo, Cálculo de estatísticas das amostras e Método TRS; além de um *plugin* no QGIS, registrado junto ao INPI. Uma vantagem destes códigos é que eles são baseados em *software* livre, podendo ser utilizados, modificados, melhorados e adaptados a diversas situações, favorecendo a pesquisa no Controle de Qualidade Cartográfica.

De modo geral, todas as hipóteses inicialmente formuladas se mostraram verdadeiras e os objetivos foram atingidos. Dessa forma, esta pesquisa pode representar um avanço, fazendo com que as linhas sejam utilizadas cada vez mais na avaliação da acurácia posicional de produtos cartográficos.

## APÊNDICES

### Capítulo 1

#### Apêndice A. Nearest Neighbor Method for Linear Features (NNMLF)

```

options(digits = 20)

# Carregar bibliotecas necessárias

library(raster)

library(sf)

library(magrittr)

library(dplyr)

library(maptools)

library(rgdal)

library(rgeos)

library(foreach)

library(doParallel)

NNMLF <- function(linhas_entrada, A){

  # Criando campo ID_Unico

  linhas_entrada$Id_Unico <- as.integer(seq(1, dim(linhas_entrada@data)[1],
by = 1))

  # Extrair os vértices das linhas de entrada

  ## Linhas e pontos possuem o mesmo "Id" com esse comando. OBS: Não é
o campo único

  vertices_linhas_entrada <- as(linhas_entrada, "SpatialPointsDataFrame")

  rownames(vertices_linhas_entrada@data) <- seq(from = 1, to =
length(vertices_linhas_entrada))

  # Preparando um data frame para receber as distâncias

```

```

tabela_dist      <-      data.frame(matrix(NA_real_,      nrow      =
dim(linhas_entrada@data)[1], ncol = dim(linhas_entrada@data)[1]))

names(tabela_dist) <- seq.int(1, dim(linhas_entrada)[1], 1)
row.names(tabela_dist) <- seq.int(1, dim(linhas_entrada)[1], 1)

# Checa quantos núcleos existem
ncl <- detectCores()

ncl

# Registra os clusters a serem utilizados
cl <- makeCluster(ncl)

registerDoParallel(cl)

# CALCULANDO A DISTANCIA - UTILIZANDO PROGRAMACAO EM
PARALELO

tabela_dist <- foreach (i = 1:dim(linhas_entrada)[1]) %:% # Pontos

  foreach (j = 1:dim(linhas_entrada)[1], .combine = 'c') %dopar%{ # Linhas

    if (i != j){ # Otimizar, para não calcular distâncias desnecessárias
mean(rgeos::gDistance(vertices_linhas_entrada[vertices_linhas_entrada@data$Id_U
nico == i,], linhas_entrada[linhas_entrada@data$Id_Unico == j,], byid = TRUE))

    } else {

      0 # Distância dos pontos para a mesma linha de origem

    }

  }

tabela_dist <- as.data.frame(tabela_dist)

colnames(tabela_dist) <- seq(1, dim(tabela_dist)[2], by = 1)

# Calcular a Distância de Hausdorff

Dist_Hausdorff      <-      data.frame(matrix(NA_real_,      nrow      =
dim(linhas_entrada@data)[1], ncol = dim(linhas_entrada@data)[1]))

names(Dist_Hausdorff) <- seq.int(1,dim(linhas_entrada)[1],1)

```

```

row.names(Dist_Hausdorff) <- seq.int(1,dim(linhas_entrada)[1],1)
for (i in 1:dim(linhas_entrada)[1]) { # Pontos
  for (j in 1:dim(linhas_entrada)[1]) { # Linhas
    if (i != j){
      if (tabela_dist[i,j] < tabela_dist[j,i]){
        Dist_Hausdorff[i,j] <- tabela_dist[j,i]
      } else {
        Dist_Hausdorff[i,j] <- tabela_dist[i,j]
      }
    } else {
      Dist_Hausdorff[i,j] <- 0
    }
  }
}

NNMLF <- foreach (i = 1:dim(linhas_entrada)[1], .combine = 'rbind') %dopar%{
  Dist_Hausdorff[i, order(Dist_Hausdorff[i,], decreasing = FALSE)[2]]
}

NNMLF_2 <- foreach (i = 1:dim(linhas_entrada)[1], .combine = 'rbind')
%dopar%{
  Dist_Hausdorff[i, order(Dist_Hausdorff[i,], decreasing = FALSE)[3]]
}

NNMLF_3 <- foreach (i = 1:dim(linhas_entrada)[1], .combine = 'rbind')
%dopar%{
  Dist_Hausdorff[i, order(Dist_Hausdorff[i,], decreasing = FALSE)[4]]
}

```

```

# R Observado
R_OBS <- sum(NNMLF)/dim(linhas_entrada@data)[1]
R_OBS_2 <- sum(NNMLF_2)/dim(linhas_entrada@data)[1]
R_OBS_3 <- sum(NNMLF_3)/dim(linhas_entrada@data)[1]

# R Esperado
Gamma_1 <- c(0.5000, 0.7500, 0.9375, 1.0937, 1.2305, 1.3535)
Gamma_2 <- c(0.2613, 0.2722, 0.2757, 0.2775, 0.2784, 0.2789)
Gamma <- data.frame(Gamma_1, Gamma_2)

R_ESP <- Gamma[1,1]*sqrt(A/dim(linhas_entrada@data)[1])
R_ESP_2 <- Gamma[2,1]*sqrt(A/dim(linhas_entrada@data)[1])
R_ESP_3 <- Gamma[3,1]*sqrt(A/dim(linhas_entrada@data)[1])

# Calcular a estatística R
R <- R_OBS/R_ESP
R_2 <- R_OBS_2/R_ESP_2
R_3 <- R_OBS_3/R_ESP_3

# Aplicar o teste Z
SE <- Gamma[1,2]*sqrt(A/(dim(linhas_entrada@data)[1])^2)
SE_2 <- Gamma[2,2]*sqrt(A/(dim(linhas_entrada@data)[1])^2)
SE_3 <- Gamma[3,2]*sqrt(A/(dim(linhas_entrada@data)[1])^2)

Z <- (R_OBS - R_ESP)/SE
Z_2 <- (R_OBS_2 - R_ESP_2)/SE_2

```

```
Z_3 <- (R_OBS_3 - R_ESP_3)/SE_3

# Considerando um nível de confiança de 95% (nível de significância de 5%),
temos:

Z_tab <- 1.96

# 1 - AGRUPADO
# 2 - ALEATORIO
# 3 - DISPERSO

# NA - NÚMERO DE LINHAS INSUFICIENTE PARA A ORDEM EM
QUESTÃO

# Inferir sobre o padrão de distribuição espacial das linhas
# Primeira Ordem

if (dim(linhas_entrada@data)[1] > 1){
  if (abs(Z) < Z_tab){ # Hipotese Nula não rejeitada
    indice_NNMLF <- 2
  } else {
    if (R < 1){
      indice_NNMLF <- 1
    } else {
      indice_NNMLF <- 3
    }
  }
} else {
  indice_NNMLF <- NA
}
```

```
# Segunda Ordem
if (dim(linhas_entrada@data)[1] > 2){
  if (abs(Z_2) < Z_tab){ # Hipotese Nula não rejeitada
    indice_NNMLF_2 <- 2
  } else {
    if (R_2 < 1){
      indice_NNMLF_2 <- 1
    } else {
      indice_NNMLF_2 <- 3
    }
  }
} else {
  indice_NNMLF_2 <- NA
}

# Terceira Ordem
if (dim(linhas_entrada@data)[1] > 3){
  if (abs(Z_3) < Z_tab){ # Hipotese Nula não rejeitada
    indice_NNMLF_3 <- 2
  } else {
    if (R_3 < 1){
      indice_NNMLF_3 <- 1
    } else {
      indice_NNMLF_3 <- 3
    }
  }
} else {
```

```
    indice_NNMLF_3 <- NA
  }
  # Stop clusters
  stopCluster(cl)

  saida_NNMLF <- cbind(indice_NNMLF, indice_NNMLF_2, indice_NNMLF_3)
  return(saida_NNMLF)
}
```

## Capítulo 2

### Apêndice A. *Buffer Simples*

```

# Esta é uma função auxiliar do método Buffer Simples
Buffer_Simples <- function(linhas_referencia, linhas_teste, largura_buffer){
  # Adicionando o campo ID_Unico_2
  linhas_referencia <- ID_Unico_Seq(linhas_referencia)
  linhas_teste <- ID_Unico_Seq(linhas_teste)
  # Calcular o comprimento das linhas de teste, em metros
  linhas_teste@data$Compri_calc_m <- gLength(linhas_teste, byid = TRUE)
  # Aplicar um buffer, de comprimento 'largura_buffer', nas linhas de referência
  buffer_ref <- gBuffer(linhas_referencia, byid = TRUE, capStyle = "round", width
= largura_buffer, quadsegs = 25)
  # Criando a tabela que irá retornar o resultado do método do Buffer Simples
  tabela_buffer_simples <- linhas_teste@data
  # Criando as variáveis para serem utilizadas na obtenção do resultado
  tabela_buffer_simples$larg_buff <- rep(NA, dim(linhas_teste)[1])
  tabela_buffer_simples$CLX <- rep(NA, dim(linhas_teste)[1])
  tabela_buffer_simples$px <- rep(NA, dim(linhas_teste)[1])
  for(i in 1:dim(linhas_referencia)[1]){
    # Obtendo a largura do buffer
    tabela_buffer_simples[tabela_buffer_simples$ID_Unico_2 == i, ]$larg_buff
<- largura_buffer
    # Realizar a interseção entre as linhas de teste e o buffer das linhas de
referência
    intersec_buffRef_teste <-
gIntersection(buffer_ref[tabela_buffer_ref@data$ID_Unico_2 == i, ],
linhas_teste[linhas_teste@data$ID_Unico_2 == i, ])

```

```

        if(!is.null(intersec_buffRef_teste)){ # Teste se há interseção entre o buffer e
a linha

            # Manipulação para obter a soma do comprimento, em metros, para cada
linha de teste resultante

            # da interseção

            ID_linhas_intersec                <-                seq.int(1,
length(intersec_buffRef_teste@lines[[1]]@Lines), 1)

            ID_linhas_intersec <- as.data.frame(ID_linhas_intersec)

            intersec_buffRef_teste.df                <-
SpatialLinesDataFrame(intersec_buffRef_teste, ID_linhas_intersec, match.ID =
TRUE)

            soma_compri_m_intersec <- gLength(intersec_buffRef_teste.df) # CLX

            # Salvar o comprimento da interseção (soma_compri_m_intersec = CLX)
na tebla do resultado do método

            tabela_buffer_simples[tabela_buffer_simples$ID_Unico_2 == i,]$CLX <-
soma_compri_m_intersec

            # Cálculo do percentual de inclusão (p_incl) de cada linha teste dentro do
buffer de referência

            tabela_buffer_simples[tabela_buffer_simples$ID_Unico_2 == i,]$px <-
tabela_buffer_simples[tabela_buffer_simples$ID_Unico_2 == i,]$CLX/

            tabela_buffer_simples[tabela_buffer_simples$ID_Unico_2                ==
i,]$Compri_calc_m

        } else {

            tabela_buffer_simples[tabela_buffer_simples$ID_Unico_2 == i,]$px <- 0

        }

    }

    # Retorna o resultado

    return(tabela_buffer_simples)

```

```

}

# Esta é a função principal do método Buffer Simples
Buffer_Simples_Original <- function(linhas_referencia, linhas_teste,
tabela_DH){
  # Calculando o passo da largura do buffer para cada linha (valores diferentes)
  tabela_DH$passo_buffer <- (tabela_DH$DH)/20 # Fazer 20 iterações
  # Adicionando o campo ID_Unico_2
  linhas_referencia <- ID_Unico_Seq(linhas_referencia)
  linhas_teste <- ID_Unico_Seq(linhas_teste)
  # Criando a tabela que irá retornar o resultado do método do Buffer Simples
  tabela_buffer_simples_Final <- linhas_teste@data
  # Criando a variável para serem utilizadas na obtenção do resultado
  tabela_buffer_simples_Final$discrep <- rep(NA, dim(linhas_teste)[1])
  list_buffer_simples_final <- list()
  for (i in 1:dim(linhas_referencia)[1]) {
    px_min <- 0
    k <- 1
    list_buffer_simples <- list()
    while(px_min < 0.91){
      # Determinar a largura do buffer a cada iteração
      largura_buffer <- (tabela_DH[tabela_DH$ID_Unico_2 ==
i,]$passo_buffer)*k
      # Aplicar a função do Buffer Simples feita anteriormente para cada largura
de buffer
      tabela_buffer_simples <-
Buffer_Simples(linhas_referencia[linhas_referencia$ID_Unico_2 == i,],

```

```

        linhas_teste[linhas_teste$ID_Unico_2 == i,]
        largura_buffer)

    list_buffer_simples[[k]] <- tabela_buffer_simples

    px_min <- min(tabela_buffer_simples$px)

    k <- k + 1

    print(largura_buffer)

    print(px_min)

    print(tabela_buffer_simples[tabela_buffer_simples$px
px_min,]$ID_Unico)

}

# Transformar a saída em data frame

list_buffer_simples.df <- do.call(rbind.data.frame, list_buffer_simples)

# Armazena a lista final com os resultados

list_buffer_simples_final[[i]] <- list_buffer_simples.df

}

# Transformar a saída da lista FINAL em data frame

list_buffer_simples_final.df <- do.call(rbind.data.frame,
list_buffer_simples_final)

# Atualizando o campo ID_Unico2

list_buffer_simples_final.df$ID_Unico_2 <-
dplyr::dense_rank(list_buffer_simples_final.df$ID_Unico)

# Preparando variáveis

list_buffer_simples.df <- list_buffer_simples_final.df

# Checa quantos núcleos existem e Registra os clusters a serem utilizados

ncl<-detectCores()

cl <- makeCluster(ncl)

registerDoParallel(cl)

```

```

tab_buffer_simples_final <- foreach (i = 1:dim(linhas_referencia)[1], .combine
= 'rbind') %dopar% {

  list_buffer_simples.df_selec <-
list_buffer_simples.df[list_buffer_simples.df$ID_Unico_2 == i,]

  vetor_90_perc <- rep(0.900, dim(list_buffer_simples.df_selec)[1])

  linha_curva <- sp::Line(matrix(c(list_buffer_simples.df_selec$larg_buff,
list_buffer_simples.df_selec$px), nc = 2, byrow = F))

  linha_90_perc <- sp::Line(matrix(c(list_buffer_simples.df_selec$larg_buff,
vetor_90_perc), nc = 2, byrow = F))

  linha_curva_1 <- sp::Lines(list(linha_curva), ID = "1")
  linha_90_perc_1 <- sp::Lines(list(linha_90_perc), ID = "1")

  linha_curva_sl <- sp::SpatialLines(list(linha_curva_1), proj4string =
sp::CRS(linhas_referencia@proj4string@projargs))

  linha_90_perc_sl <- sp::SpatialLines(list(linha_90_perc_1), proj4string =
sp::CRS(linhas_referencia@proj4string@projargs))

  # Verifica se há interseção

  if(dim(linha_curva_sl@lines[[1]]@Lines[[1]]@coords)[1] < 2 &
  linha_curva_sl@lines[[1]]@Lines[[1]]@coords[2] > 0.90){

    int.coords <- matrix(data = NA, nrow = 1, ncol = 2)

    int.coords[1] <- linha_curva_sl@lines[[1]]@Lines[[1]]@coords[2]
    int.coords[2] <- linha_curva_sl@lines[[1]]@Lines[[1]]@coords[2]

  } else {

    # Calcula onde as linhas se intersectam

    int.pts <- rgeos::gIntersection(linha_curva_sl, linha_90_perc_sl, byid =
TRUE)

    int.coords <- int.pts@coords

  }
}

```

```

# Armazenar o resultado

tabela_buffer_simples_Final$ID_Unico_2 <-
dplyr::dense_rank(tabela_buffer_simples_Final$ID_Unico)

tabela_buffer_simples_Final[tabela_buffer_simples_Final$ID_Unico_2 ==
i,]$discrep <- int.coords[1]

# Preparar os vetores para comporem a tabela final do resultado

ID_Unico <-
tabela_buffer_simples_Final[tabela_buffer_simples_Final$ID_Unico_2 == i,
]$ID_Unico

ID_Unico_2 <-
tabela_buffer_simples_Final[tabela_buffer_simples_Final$ID_Unico_2 == i,
]$ID_Unico_2

discrep <-
tabela_buffer_simples_Final[tabela_buffer_simples_Final$ID_Unico_2 == i,]$discrep

return(data.frame(ID_Unico, ID_Unico_2, discrep))
}

# Stop clusters

stopCluster(cl)

# Retorna o resultado do método do Buffer Simples Original

return(tab_buffer_simples_final)
}

```

## Apêndice B. *Buffer Duplo*

```

# Esta é uma função auxiliar do método Buffer Duplo

Buffer_Duplo <- function(linhas_referencia, linhas_teste, largura_buffer){

  # Adicionando o campo ID_Unico_2

  linhas_referencia <- ID_Unico_Seq(linhas_referencia)

  linhas_teste <- ID_Unico_Seq(linhas_teste)

  # Aplicando um buffer de largura 'largura_buffer' nas linhas de teste e de
referência

  buffer_ref <- gBuffer(linhas_referencia, byid = TRUE, capStyle = "round", width
= largura_buffer, quadsegs = 25)

  buffer_teste <- gBuffer(linhas_teste, byid = TRUE, capStyle = "round", width =
largura_buffer, quadsegs = 25)

  # Criando a tabela que irá retornar o resultado do método do Buffer Duplo

  tabela_buffer_duplo <- buffer_teste@data

  # Criando as variáveis para serem utilizadas na obtenção do resultado

  tabela_buffer_duplo$larg_buff <- rep(NA, dim(buffer_teste)[1])

  tabela_buffer_duplo$AT <- rep(NA, dim(buffer_teste)[1])

  tabela_buffer_duplo$AF <- rep(NA, dim(buffer_teste)[1])

  tabela_buffer_duplo$dm <- rep(NA, dim(buffer_teste)[1])

  for (i in 1:dim(buffer_teste)[1]){

    # Obtendo a área do polígono do buffer das linhas de teste

    tabela_buffer_duplo[tabela_buffer_duplo$ID_Unico_2 == i, ]$AT <-
buffer_teste[buffer_teste@data$ID_Unico_2 == i, ]@polygons[[1]]@area

    # Obtendo a largura do buffer

    tabela_buffer_duplo[tabela_buffer_duplo$ID_Unico_2 == i, ]$larg_buff <-
largura_buffer

```

```

# Realizando a interseção entre os buffers de teste e de referência
diff_ref_teste <- gDifference(buffer_ref[buffer_ref@data$ID_Unico_2 == i, ],
                             buffer_teste[buffer_teste@data$ID_Unico_2 == i, ])

# Obtendo a área da interseção, com uma condição para o caso em que as
linhas de teste e de

# referência forem iguais
if (is.null(diff_ref_teste)){
  tabela_buffer_duplo[tabela_buffer_duplo$ID_Unico_2 == i, ]$AF <- 0
} else {
  tabela_buffer_duplo[tabela_buffer_duplo$ID_Unico_2 == i, ]$AF <-
diff_ref_teste@polygons[[1]]@area
}

# Obtendo a discrepância posicional do método do Buffer Duplo
tabela_buffer_duplo[tabela_buffer_duplo$ID_Unico_2 == i, ]$dm <-
(pi*largura_buffer*tabela_buffer_duplo[tabela_buffer_duplo$ID_Unico_2 == i,
]$AF)/tabela_buffer_duplo[tabela_buffer_duplo$ID_Unico_2 == i, ]$AT
}

# Retorna o resultado
return(tabela_buffer_duplo)
}

# Esta é a função principal do método Buffer Duplo
Buffer_Duplo_Original <- function(linhas_referencia, linhas_teste, tabela_DH){
  # Calculando o passo da largura do buffer para cada linha (valores diferentes)
  tabela_DH$passo_buffer <- (tabela_DH$DH)/100 # Fazer 100 iterações

  # Adicionando o campo ID_Unico_2
  linhas_referencia <- ID_Unico_Seq(linhas_referencia)

```

```

linhas_teste <- ID_Unico_Seq(linhas_teste)
list_resultado_Areas <- list()
for(i in 1:dim(linhas_teste)[1]){
  # Criando as variáveis para serem utilizadas na obtenção do resultado
  Indice_i_linha <- vector()
  Indice_j_buffer <- vector()
  ID_Unico <- vector()
  ID_Unico_2 <- vector()
  larg_buff <- vector()
  Area_b <- vector()
  Area_c <- vector()
  Area_d <- vector()
  Area_b_n <- vector()
  Area_c_n <- vector()
  Area_d_n <- vector()
  # Seleciona a linha para cada iteração
  linhas_referencia_selec <- linhas_referencia[linhas_referencia$ID_Unico_2
== i, ]
  linhas_teste_selec <- linhas_teste[linhas_teste$ID_Unico_2 == i, ]
  for(j in 1:1000){
    largura_buffer <- (tabela_DH[tabela_DH$ID_Unico_2 == i,]$passo_buffer)*j
    # Aplicando um buffer de largura 'largura_buffer' nas linhas de teste e de
referência
    buffer_ref <- rgeos::gBuffer(linhas_referencia_selec,
                                byid = TRUE, capStyle = "round", width = largura_buffer,
quadsegs = 25)
    buffer_teste <- rgeos::gBuffer(linhas_teste_selec,

```

```

                                byid = TRUE, capStyle = "round", width = largura_buffer,
quadsegs = 25)

# Obtendo a largura do buffer, os IDs e os índices
ID_Unico[j] <- linhas_referencia_selec$ID_Unico
ID_Unico_2[j] <- linhas_teste_selec$ID_Unico_2
larg_buff[j] <- largura_buffer

Indice_i_linha[j] <- i
Indice_j_buffer[j] <- j

# Realizando os três tipos de áreas possíveis entre os buffers de teste e de
referência

if(rgeos::gIntersects(buffer_ref, buffer_teste)){

  # Area_b

  diff_ref_teste <- rgeos::gDifference(buffer_ref, buffer_teste)
  Area_b[j] <- diff_ref_teste@polygons[[1]]@area
  Area_b <- Area_b[j]

  # Area_c

  diff_teste_ref <- rgeos::gDifference(buffer_teste, buffer_ref)
  Area_c[j] <- diff_teste_ref@polygons[[1]]@area
  Area_c <- Area_c[j]

  # Area_d

  diff_intersec <- rgeos::gIntersection(buffer_teste, buffer_ref)
  Area_d[j] <- diff_intersec@polygons[[1]]@area
  Area_d <- Area_d[j]

```

```

# Realizando os três tipos de áreas possíveis - NORMALIZADAS
normalizador <- Area_b + Area_c + Area_d
Area_b_n[j] <- Area_b/normalizador
Area_c_n[j] <- Area_c/normalizador
Area_d_n[j] <- Area_d/normalizador
} else {
  Area_b[j] <- NA
  Area_c[j] <- NA
  Area_d[j] <- NA
  Area_b_n[j] <- NA
  Area_c_n[j] <- NA
  Area_d_n[j] <- NA
}
if(!is.na(Area_b_n[j]) & !is.na(Area_c_n[j]) & !is.na(Area_d_n[j])){
  if((Area_d_n[j] > Area_b_n[j]) & (Area_d_n[j] > Area_c_n[j])){
    print(unique(ID_Unico))
    break
  }
}
}
Areas_Buffer_Duplo <- data.frame(ID_Unico, ID_Unico_2, Indice_i_linha,
Indice_j_buffer,
                                larg_buff, Area_b_n, Area_c_n, Area_d_n)
list_resultado_Areas[[i]] <- Areas_Buffer_Duplo
}

```

```

list_resultado_Areas.df <- do.call(rbind.data.frame, list_resultado_Areas)
list_resultado_Areas.df <- na.omit(list_resultado_Areas.df)

list_resultado_Areas.df <-
list_resultado_Areas.df[!duplicated(list_resultado_Areas.df$Area_b_n), ]

list_resultado.df <- list_resultado_Areas.df

# Criando a tabela FINAL que irá retornar o resultado do método do Buffer
Duplo

tabela_buffer_duplo_final <- linhas_teste@data
tabela_buffer_duplo_final$buff_ideal <- rep(NA, dim(linhas_teste)[1])

# Checa quantos núcleos existem e Registra os clusters a serem utilizados
ncl<-detectCores()

cl <- makeCluster(ncl)

registerDoParallel(cl)

tabela_buffer_duplo_final <- foreach (i = 1:dim(linhas_teste)[1], .combine =
'rbind') %dopar% {

  # for(i in 1:dim(linhas_teste)[1]){

  list_resultado.df_selec <- list_resultado.df[list_resultado.df$ID_Unico_2 == i,]

  linha_b      <-      sp::Line(matrix(c(list_resultado.df_selec$larg_buff,
list_resultado.df_selec$Area_b_n), nc = 2, byrow = F))

  linha_c      <-      sp::Line(matrix(c(list_resultado.df_selec$larg_buff,
list_resultado.df_selec$Area_c_n), nc = 2, byrow = F))

  linha_d      <-      sp::Line(matrix(c(list_resultado.df_selec$larg_buff,
list_resultado.df_selec$Area_d_n), nc = 2, byrow = F))

  linha_b_1 <- sp::Lines(list(linha_b), ID = "1")
  linha_c_1 <- sp::Lines(list(linha_c), ID = "1")
  linha_d_1 <- sp::Lines(list(linha_d), ID = "1")

```

```

linha_b_sl <- sp::SpatialLines(list(linha_b_1), proj4string =
sp::CRS(linhas_referencia@proj4string@projargs))

linha_c_sl <- sp::SpatialLines(list(linha_c_1), proj4string =
sp::CRS(linhas_referencia@proj4string@projargs))

linha_d_sl <- sp::SpatialLines(list(linha_d_1), proj4string =
sp::CRS(linhas_referencia@proj4string@projargs))

# Calculate locations where spatial lines intersect

if((rgeos::gIntersects(linha_c_sl, linha_d_sl)) &
(rgeos::gIntersects(linha_b_sl, linha_d_sl))){

  int.pts <- rgeos::gIntersection(linha_c_sl, linha_d_sl, byid = TRUE)
  int.coords <- int.pts@coords

  int.pts_1 <- rgeos::gIntersection(linha_b_sl, linha_d_sl, byid = TRUE)
  int.coords_1 <- int.pts_1@coords

} else {

  if((rgeos::gIntersects(linha_c_sl, linha_d_sl)) {

    int.pts <- rgeos::gIntersection(linha_c_sl, linha_d_sl, byid = TRUE)
    int.coords <- int.pts@coords
    int.coords_1 <- int.coords

  } else {

    int.pts_1 <- rgeos::gIntersection(linha_b_sl, linha_d_sl, byid = TRUE)
    int.coords_1 <- int.pts_1@coords
    int.coords <- int.coords_1

  }

}

```

```

# Pegar o maior valor para ser a largura ideal do buffer duplo (caso mais
restritivo)

  if (int.coords[1] >= int.coords_1[1]){

    tabela_buffer_duplo_final[tabela_buffer_duplo_final$ID_Unico_2 ==
i,]$buff_ideal <- int.coords[1]

  } else{

    tabela_buffer_duplo_final[tabela_buffer_duplo_final$ID_Unico_2 ==
i,]$buff_ideal <- int.coords_1[1]

  }

  ID_Unico <-
tabela_buffer_duplo_final[tabela_buffer_duplo_final$ID_Unico_2 == i, ]$ID_Unico

  ID_Unico_2 <-
tabela_buffer_duplo_final[tabela_buffer_duplo_final$ID_Unico_2 == i, ]$ID_Unico_2

  buff_ideal <-
tabela_buffer_duplo_final[tabela_buffer_duplo_final$ID_Unico_2 == i, ]$buff_ideal

  return(data.frame(ID_Unico, ID_Unico_2, buff_ideal))

}

# Stop clusters

stopCluster(cl)

# Após a determinação da largura ideal do Buffer Duplo, calcular as
discrepâncias com esta largura

saida_Buffer_Duplo <- list()

for(i in 1:dim(linhas_teste)[1]){

  saida_Buffer_Duplo[[i]] <-
Buffer_Duplo(linhas_referencia[linhas_referencia$ID_Unico_2 == i,],
              linhas_teste[linhas_teste$ID_Unico_2 == i,],

```

```
tabela_buffer_duplo_final[tabela_buffer_duplo_final$ID_Unico_2 == i,]$buff_ideal)
  }
  Buffer_Duplo.df <- do.call(rbind.data.frame, saida_Buffer_Duplo)
  return(Buffer_Duplo.df)
}
```

## Apêndice C. Distância de Hausdorff

```

Distancia_Hausdorff <- function(linhas_referencia, linhas_teste){
  # Extrair os vértices das linhas de referência e de teste
  vertices_linhas_referencia <- as(linhas_referencia,
"SpatialPointsDataFrame")
  rownames(vertices_linhas_referencia@data) <- seq(from = 1, to =
length(vertices_linhas_referencia))
  vertices_linhas_teste <- as(linhas_teste, "SpatialPointsDataFrame")
  rownames(vertices_linhas_teste@data) <- seq(from = 1, to =
length(vertices_linhas_teste))
  # Adicionando o campo ID_Unico_2
  linhas_referencia <- ID_Unico_Seq(linhas_referencia)
  linhas_teste <- ID_Unico_Seq(linhas_teste)
  vertices_linhas_referencia <- ID_Unico_Seq(vertices_linhas_referencia)
  vertices_linhas_teste <- ID_Unico_Seq(vertices_linhas_teste)
  # Checa quantos núcleos existem e Registra os clusters a serem utilizados
  ncl<-detectCores()
  cl <- makeCluster(ncl)
  registerDoParallel(cl)
  # CALCULANDO A DISTANCIA - UTILIZANDO PROGRAMACAO EM
PARALELO
  tabela_dist_teste_ref <- foreach (i = 1:dim(linhas_teste)[1], .combine = 'rbind')
%dopar% { # %:% # Linhas
  d1 <-
mean(rgeos::gDistance(vertices_linhas_teste[vertices_linhas_teste@data$ID_Unico
_2 == i,], linhas_referencia[linhas_referencia@data$ID_Unico_2 == i,], byid = TRUE))
  ID_Unico_2 <- i

```

```

        ID_Unico <- linhas_referencia[linhas_referencia@data$ID_Unico_2 ==
i,]$ID_Unico

        return(list(ID_Unico, ID_Unico_2, d1))
    }

    # Padronizando nomes de linhas e colunas

    rownames(tabela_dist_teste_ref) <- seq(from = 1, to =
dim(tabela_dist_teste_ref)[1])

    tabela_dist_teste_ref <- as.data.frame(tabela_dist_teste_ref)

    colnames(tabela_dist_teste_ref) <- c("ID_Unico", "ID_Unico_2", "d1")

    # CALCULANDO A DISTANCIA - UTILIZANDO PROGRAMACAO EM
PARALELO

    tabela_dist_ref_teste <- foreach (i = 1:dim(linhas_referencia)[1], .combine =
'rbind') %dopar%{ # %:% # Linhas

        d2 <-
mean(rgeos::gDistance(vertices_linhas_referencia[vertices_linhas_referencia@data$
ID_Unico_2 == i,], linhas_teste[linhas_teste@data$ID_Unico_2 == i,], byid = TRUE))

        ID_Unico_2 <- i

        ID_Unico <- linhas_referencia[linhas_referencia@data$ID_Unico_2 ==
i,]$ID_Unico

        return(list(ID_Unico, ID_Unico_2, d2))
    }

    # Padronizando nomes de linhas e colunas

    rownames(tabela_dist_ref_teste) <- seq(from = 1, to =
dim(tabela_dist_ref_teste)[1])

    tabela_dist_ref_teste <- as.data.frame(tabela_dist_ref_teste)

    colnames(tabela_dist_ref_teste) <- c("ID_Unico", "ID_Unico_2", "d2")

    # Stop clusters

    stopCluster(cl)

```

```
# Unindo as tabelas das distâncias de teste e de referência e criando um
campo para DH

tabela_dist_hausdorff      <-      data.frame(tabela_dist_ref_teste,
tabela_dist_teste_ref)

tabela_dist_hausdorff$DH <- rep(NA, dim(tabela_dist_hausdorff)[1])

# Obter a Distância de Hausdorff

tabela_dist_hausdorff_dist <- tabela_dist_hausdorff[, c(3,6)]

tabela_dist_hausdorff_dist$d2 <- as.numeric(tabela_dist_hausdorff_dist$d2)

tabela_dist_hausdorff_dist$d1 <- as.numeric(tabela_dist_hausdorff_dist$d1)

tabela_dist_hausdorff$DH <- apply(X = tabela_dist_hausdorff_dist, MARGIN
= 1, FUN = max)

# Retorna o resultado

return(tabela_dist_hausdorff)

}
```

## Apêndice D. Influência do Vértice

```

Influencia_Vertice <- function(linhas_referencia, linhas_teste){
  # Adicionando o campo ID_Unico_2
  linhas_referencia <- ID_Unico_Seq(linhas_referencia)
  linhas_teste <- ID_Unico_Seq(linhas_teste)

  # Calcular o comprimento (CLR) e extrair os vértices de cada linha i de
  REFERÊNCIA

  vertices_linhas_referencia <- as(linhas_referencia,
  "SpatialPointsDataFrame")

  rownames(vertices_linhas_referencia@data) <- seq(from = 1, to =
  length(vertices_linhas_referencia))

  linhas_referencia@data$CLR <- gLength(linhas_referencia, byid = TRUE)

  # Calcular a menor distância (dhk) para cada vértice k da linha referência em
  relação a

  # qualquer segmento da linha teste.

  # Criar um campo sequencial antes, para identificar cada vértice das linhas de
  referência

  vertices_linhas_referencia$seq <- seq.int(from = 1, to =
  length(vertices_linhas_referencia), by = 1)

  list_dhk <- list()

  list_ID_dhk <- list()

  for(i in 1:dim(linhas_teste)[1]){

    list_dhk[[i]] <-
  rgeos::gDistance(vertices_linhas_referencia[vertices_linhas_referencia$ID_Unico_2
  == i,], linhas_teste[linhas_teste$ID_Unico_2 == i,], byid = TRUE)

    list_ID_dhk[[i]] <-
  vertices_linhas_referencia[vertices_linhas_referencia$ID_Unico_2 == i,]$seq

```

```

}

# Converter as listas para data frames

list_dhk.unlist <- unlist(list_dhk, recursive = TRUE, use.names = TRUE)

list_ID_dhk.unlist <- unlist(list_ID_dhk, recursive = TRUE, use.names = TRUE)

dhk.df <- data.frame(list_ID_dhk.unlist, list_dhk.unlist)

names(dhk.df) <- c("seq", "dhk")

# Unir o data frame das menores distâncias com a tabela de atributos do dado
espacial

vertices_linhas_referencia <- sp::merge(vertices_linhas_referencia, dhk.df, by
= c("seq", "seq"))

# Transferir a soma dos segmentos de retas adjacentes aos vértices para a
tabela dos vértices

tabela_inf_vert <- vertices_linhas_referencia@data

teste_st_join_buff_list <- list()

for(i in 1:dim(linhas_teste)[1]){

  # Obter os segmentos de reta

  seg_linhas_referencia <-
eixport::st_explode(linhas_referencia[linhas_referencia$ID_Unico_2 == i,])

  seg_linhas_referencia$CLR_seg <-
gLength(as_Spatial(seg_linhas_referencia), byid = TRUE)

  # Aplicar um buffer de 10 cm para fazer a junção espacial com os segmentos
de reta

  vertices_linhas_referencia_buff <-
st_buffer(st_as_sf(vertices_linhas_referencia[vertices_linhas_referencia$ID_Unico_2
== i,]), dist = 0.10)

  teste_st_join_buff <- st_join(st_as_sf(vertices_linhas_referencia_buff),
seg_linhas_referencia)

```

```

# Retirar a geometria do buffer, que não é mais necessária
teste_st_join_buff <- st_drop_geometry(teste_st_join_buff)

# Deixar somente as colunas essenciais
teste_st_join_buff <- subset(teste_st_join_buff, select = c('seq','CLR_seg'))

# Guardar o resultado em uma list
teste_st_join_buff_list[[i]] <- teste_st_join_buff
}

teste_st_join_buff_df <- do.call(rbind.data.frame, teste_st_join_buff_list)

seq <- vector()

CLR_seg <- vector()

for (i in 1:length(vertices_linhas_referencia)){
  seq[i] <- i
  CLR_seg[i] <- sum(teste_st_join_buff_df[teste_st_join_buff_df$seq ==
i,]$CLR_seg)
}

df_inf_vert_prev <- data.frame(seq, CLR_seg)

df_inf_vert <- sp::merge(tabela_inf_vert, df_inf_vert_prev, by = c("seq", "seq"))

# Calcular o denominador da fórmula da discrepância (Di)
df_inf_vert$denom <- (df_inf_vert$dhk)*(df_inf_vert$CLR_seg)

# Calcular a SOMA do denominador da fórmula da discrepância (Di), para
cada linha e

# Gerar a tabela final do Método Influência do Vértice
tabela_Influencia_Vertice <- linhas_referencia@data

tabela_Influencia_Vertice$denom_sum <- rep(NA,
dim(tabela_Influencia_Vertice)[1])

for(i in 1:length(linhas_referencia)){

```

```
tabela_Influencia_Vertice[tabela_Influencia_Vertice$ID_Unico_2 ==  
i,]$denom_sum <- sum(df_inf_vert[df_inf_vert$ID_Unico_2 == i,]$denom)  
}  
tabela_Influencia_Vertice$Di <-  
(tabela_Influencia_Vertice$denom_sum)/(2*tabela_Influencia_Vertice$CLR)  
# Retorna o resultado  
return(tabela_Influencia_Vertice)  
}
```

## Apêndice E. Seleção da Amostra

```

Selecao_Amostra <- function(feicoes_janela, tamanho_amostral_m, tol_porc){
  # Para garantir que é um objeto 'sf' (Facilita para a exclusão dos NA's)
  feicoes_janela <- st_as_sf(feicoes_janela)

  # Ordenar para aumentar as chances do padrão do NNMLF ser disperso ou
  aleatório

  feicoes_janela <- feicoes_janela[order(feicoes_janela$Compri_m,
                                         decreasing = TRUE),]

  # Preparando para a seleção
  utilizar <- rep(NA, dim(feicoes_janela)[1])
  feicoes_janela$utilizar <- utilizar
  soma_comprimento <- 0
  for (i in 1:dim(feicoes_janela)[1]){
    if(feicoes_janela$Compri_m[i] > (tamanho_amostral_m)*(1 +
(tol_porc/100))){
      feicoes_janela$utilizar[i] <- NA
    } else {
      if ((soma_comprimento + feicoes_janela$Compri_m[i]) <=
(tamanho_amostral_m)*(1 + (tol_porc/100))){
        soma_comprimento <- soma_comprimento + feicoes_janela$Compri_m[i]
        feicoes_janela$utilizar[i] <- 1
        if(soma_comprimento > tamanho_amostral_m){
          break
        }
      }
    }
  }
}

```

```

# Se chegou no final e não tem o tamanho amostral ainda, inverte o data frame
para tentar novamente

if((i == dim(feicoes_janela)[1]) & (soma_comprimento <
tamanho_amostral_m)){

  # Aplicar NA em toda a coluna 'utilizar'
  feicoes_janela$utilizar <- NA

  soma_comprimento <- 0

  feicoes_janela_split <- split(feicoes_janela,
rep(1:2,each=length(feicoes_janela)/2))

  feicoes_janela_split1 <- feicoes_janela_split[[1]]
  feicoes_janela_split2 <- feicoes_janela_split[[2]]

  feicoes_janela <- rbind(feicoes_janela_split2, feicoes_janela_split1)

  for (i in 1:dim(feicoes_janela)[1]){

    if(feicoes_janela$Compri_m[i] > (tamanho_amostral_m)*(1 +
(tol_porc/100))){

      feicoes_janela$utilizar[i] <- NA

    } else {

      if ((soma_comprimento + feicoes_janela$Compri_m[i]) <=
(tamanho_amostral_m)*(1 + (tol_porc/100))){

        soma_comprimento <- soma_comprimento +
feicoes_janela$Compri_m[i]

        feicoes_janela$utilizar[i] <- 1

        if(soma_comprimento > tamanho_amostral_m){

          break

        }

      }

    }

  }

}

```

```
}  
feicoes_janela <- na.omit(feicoes_janela)  
if(dim(feicoes_janela)[1] != 0){  
  if((dim(feicoes_janela)[1] == 1) | (soma_comprimento <  
tamanho_amostral_m)){  
    feicoes_janela$utilizar <- NA  
    feicoes_janela <- na.omit(feicoes_janela)  
  }  
}  
  
# Antes de retornar o resultado, transformar em SpatialDataFrame  
if(dim(feicoes_janela)[1] != 0){  
  feicoes_janela <- as_Spatial(feicoes_janela)  
}  
  
# Retorna o resultado  
return(feicoes_janela)  
}
```

## Apêndice F. Simulação Monte Carlo

```

# Armazenar o resultado da Simulação Monte Carlo em uma lista
resultado_simulacao <- list()

contador_NNMLF <- 0

contador_tamanho_amostral <- 0

# O tamanho amostral mínimo é dado pelo lado da área (Area_m2)
tamanho_amostral_min <- sqrt(Area_m2)

# Passo para o tamanho amostral = 50% do tamanho amostral mínimo
passo_tamanho_amostral <- tamanho_amostral_min*0.50

# Início da Simulação Monte Carlo
while(length(resultado_simulacao) < 10000){
  # Gera um conjunto de pontos aleatórios
  sample_points <- st_sample(st_as_sf(Area_Estudo_clip), size = 10000, # nc
= limite da Área de Estudo
                                type = 'random', exact = TRUE, by_polygon = TRUE) %>%
  # optional: Give the points an ID
  st_sf('ID' = seq(length(.)), 'geometry' = .) %>%
  # optional: Get underlying polygon attributes
  st_intersection(., st_as_sf(Area_Estudo_clip))
  linhas_referencia_sfc <- st_as_sf(linhas_referencia)
  # Criar a Janela de Busca
  polysB_sfc_selec <- st_buffer(sample_points,(sqrt(Area_m2)/2), nQuadSegs
= 5 ,
                                endCapStyle = 'SQUARE')
  # Obtenção do Centróide da Janela de Busca - Trocar 1 por k

```

```

# rgeos::gCentroid(as_Spatial(polysB_sfc_selec[1,]))@coords[1] - X
# rgeos::gCentroid(as_Spatial(polysB_sfc_selec[1,]))@coords[2] - Y

for(k in 1:dim(polysB_sfc_selec)[1]){
  # teste_sp <- as_Spatial(polysB_sfc_selec)
  # names(teste_sp@polygons) <- seq.int(1, dim(teste_sp)[1], 1)
  # teste_sp$id <- seq.int(1, dim(teste_sp)[1], 1)

  # Realizar a seleção das feições que estão completamente dentro da Janela
de Busca

  linhas_janela <- st_join(st_as_sf(linhas_referencia_sfc),
                           polysB_sfc_selec[k,], join = st_covered_by,
                           left = FALSE) # rod_intern = linhas_janela

  linhas_janela$ID_Unico <- as.integer(linhas_janela$ID_Unico)

  # salvar a área (m2) e as coordenadas do centro da Janela de Busca
  ## Somente se houverem feições selecionadas na janela
  if (dim(linhas_janela)[1] != 0){

    linhas_janela$Coord_X <-
rgeos::gCentroid(as_Spatial(polysB_sfc_selec[k,]))@coords[1] #a[[1]]@labpt[1]

    linhas_janela$Coord_Y <-
rgeos::gCentroid(as_Spatial(polysB_sfc_selec[k,]))@coords[2] #a[[1]]@labpt[2]

    linhas_janela$Area_m2 <- Area_m2 #a[[1]]@area

  }

  # Gerar uma janela de procura que irá selecionar aleatoriamente feições
dentro da área determinada

  # linhas_Janela <- Janela_Busca(linhas_referencia, Area_m2)

  if((dim(linhas_janela)[1] > 1) & (sum(linhas_janela$Compri_m) >=
tamanho_amostral_min)){

```

```

linhas_amostra      <-      Selecao_Amostra(linhas_janela,
tamanho_amostral_min, tol_porc)

# dim(linhas_amostra)[1]

# Condição para verificar se a seleção da amostra retornou alguma feição.
# Se não tiver retornado, pular esse laço do loop
if(dim(linhas_amostra)[1] < 2){
  next
}
} else {
  next # Ir direto para a próxima iteração (pular esse laço do loop)
}

# Avaliar o padrão de distribuição espacial da amostra utilizando o método
NNMLF

indice_NNMLF <- NNMLF(linhas_amostra, Area_m2)

# Verificar se o padrão de distribuição espacial da amostra é aleatório (2) ou
disperso (3)

if(indice_NNMLF[1] == 2 | indice_NNMLF[1] == 3){ # Avalia só a primeira
ordem do NNMLF

  # 1. Recuperar as discrepâncias posicionais da tabela da população pelo
ID_Unico

  discrep_amostra      <-
tabela_discrep_populacao[which(tabela_discrep_populacao$ID_Unico      %in%
linhas_amostra$ID_Unico),]

  # 2. Calcular as estatísticas básicas das discrepâncias de cada método
estatist_amostra_BS <- Estatisticas_Basicas(discrep_amostra$discrep)
estatist_amostra_BD <- Estatisticas_Basicas(discrep_amostra$dm)
estatist_amostra_DH <- Estatisticas_Basicas(discrep_amostra$DH)
estatist_amostra_IV <- Estatisticas_Basicas(discrep_amostra$Di)

```

# 3. Classificar as amostras

```
Classificacao_BS <- Classificacao(discrep_amostra$discrep,
tabela_PEC_EP, Denom_Escala)
```

```
Classificacao_BD <- Classificacao(discrep_amostra$dm, tabela_PEC_EP,
Denom_Escala)
```

```
Classificacao_DH <- Classificacao(discrep_amostra$DH, tabela_PEC_EP,
Denom_Escala)
```

```
Classificacao_IV <- Classificacao(discrep_amostra$Di, tabela_PEC_EP,
Denom_Escala)
```

# 4. Armazenar o resultado da Simulação na lista 'resultado\_simulacao'

# SALVAR DA SIMULAÇÃO:

# 1. Coordenadas do centro da Janela de Busca

```
Coord_Janela <- data.frame(linhas_amostra$Coord_X[1],
linhas_amostra$Coord_Y[1])
```

```
names(Coord_Janela) <- c("X", "Y")
```

# 2. Tamanho da Área

```
Area_m2_Janela <- linhas_amostra$Area_m2[1]
```

# 3. Tamanho da amostra (objetivo)

```
tamanho_amostrado_alvo <- tamanho_amostrado_min
```

# 4. Tamanho da amostra (utilizado)

```
tamanho_amostrado_real <- sum(linhas_amostra$Compri_m)
```

# 5. Tolerância do tamanho da amostra

```
# tol_porcentagem
```

# 6. Índice NNMLF

```
# indice_NNMLF[1]
```

```

# 7. ID_Único das feições que compõem a amostra
# discrep_amostra

# 9. Denominador da Escala utilizada para avaliação
# Denom_Escala

# 10. Estatísticas Básicas das discrepâncias
Estatisticas_Amostra <- list(estatist_amostra_BS, estatist_amostra_BD,
                             estatist_amostra_DH, estatist_amostra_IV)
names(Estatisticas_Amostra) <- c("BS", "BD", "DH", "IV")

# 11. Classificação das amostras
Classificacao_Amostra <- list(Classificacao_BS, Classificacao_BD,
                              Classificacao_DH, Classificacao_IV)
names(Classificacao_Amostra) <- c("BS", "BD", "DH", "IV")

# Salvar os resultados deste laço do Loop
resultado_simulacao_parcial <- list()
resultado_simulacao_parcial <- append(resultado_simulacao_parcial,
Coord_Janela)
resultado_simulacao_parcial <- append(resultado_simulacao_parcial,
Area_m2_Janela)
names(resultado_simulacao_parcial)[3] <- c("Area m2")
resultado_simulacao_parcial <- append(resultado_simulacao_parcial,
tamanho_amostral_alvo)
names(resultado_simulacao_parcial)[4] <- c("Tamanho Amostral Alvo")
resultado_simulacao_parcial <- append(resultado_simulacao_parcial,
tamanho_amostral_real)
names(resultado_simulacao_parcial)[5] <- c("Tamanho Amostral Utilizado")

```

```

        resultado_simulacao_parcial <- append(resultado_simulacao_parcial,
tol_porc)

        names(resultado_simulacao_parcial)[6] <- c("Tolerância Porc")

        resultado_simulacao_parcial <- append(resultado_simulacao_parcial,
indice_NNMLF[1])

        names(resultado_simulacao_parcial)[7] <- c("Índice NNMLF")

        resultado_simulacao_parcial[[8]] <- discrep_amostra

        names(resultado_simulacao_parcial)[8] <- c("Discrepâncias Amostra")

        resultado_simulacao_parcial <- append(resultado_simulacao_parcial,
Denom_Escala)

        names(resultado_simulacao_parcial)[9] <- c("Denominador da Escala")

        resultado_simulacao_parcial[[10]] <- Estatisticas_Amostra

        names(resultado_simulacao_parcial)[10] <- c("Estatísticas Amostra")

        resultado_simulacao_parcial[[11]] <- Classificacao_Amostra

        names(resultado_simulacao_parcial)[11] <- c("Classificação Amostra")

        # Salvando no data frame final

        # resultado_simulacao[[length(resultado_simulacao) + 1]] <-
resultado_simulacao_parcial

        # resultado_simulacao <- append(resultado_simulacao,
resultado_simulacao_parcial)#, after = length(resultado_simulacao))

        resultado_simulacao[[length(resultado_simulacao) + 1]] <-
resultado_simulacao_parcial

        if(length(resultado_simulacao) == 10000){

            break

        }

        contador_NNMLF <- 0

        contador_tamanho_amostral <- contador_tamanho_amostral + 1

        print(contador_tamanho_amostral)

```

```
        if(contador_tamanho_amostral == 1000){ # 1000 amostras para cada
tamanho amostral

            tamanho_amostral_min <- tamanho_amostral_min +
passo_tamanho_amostral

            contador_tamanho_amostral <- 0

            save.image(file = 'Resultado_Simulacao.Rdata')

        }

    } else{

        contador_NNMLF <- contador_NNMLF + 1

    }

}

}
```

## Apêndice G. Estatísticas das Amostras

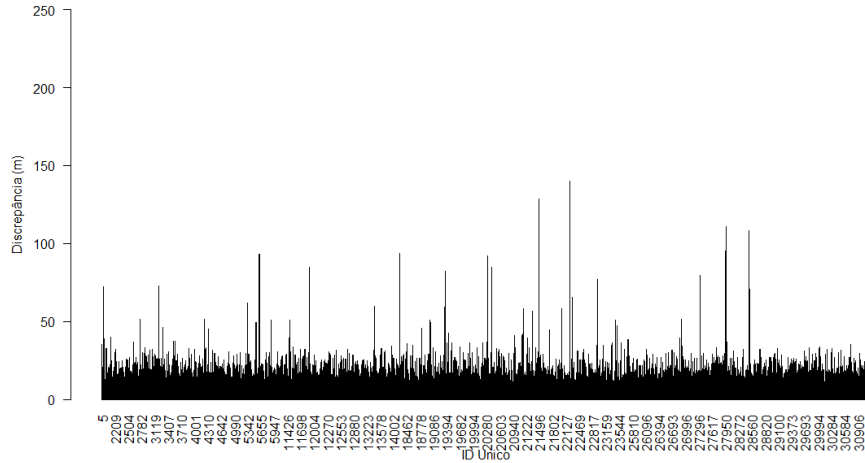
```

Estatisticas_Basicas <- function(vetor_discrepancias){
  # Obter o RMS das discrepâncias posicionais
  RMS <-
sqrt((sum(vetor_discrepancias*vetor_discrepancias))/(length(vetor_discrepancias) -
1))

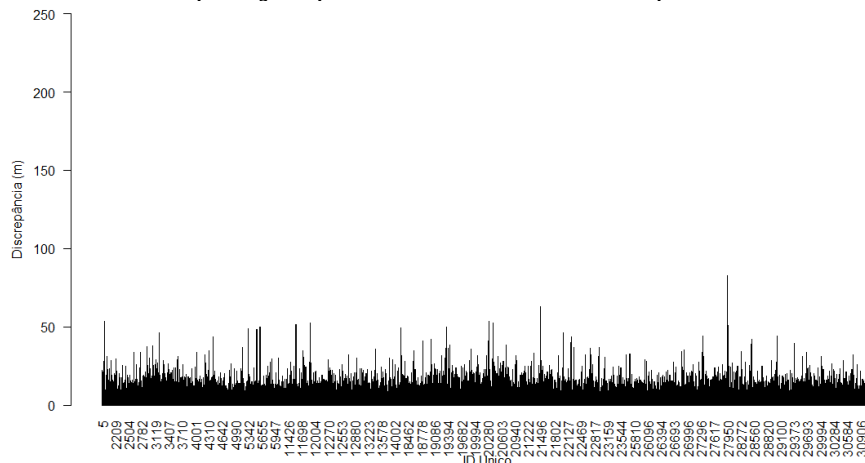
  # Obter a Média das discrepâncias posicionais
  Media <- mean(vetor_discrepancias)
  # Obter a Mediana das discrepâncias posicionais
  Mediana <- median(vetor_discrepancias)
  # Obter o Desvio-Padrão das discrepâncias posicionais
  Desvio_Padrao <- sd(vetor_discrepancias)
  # Preparando a saída dos dados, em formato de data frame
  saida_Estatisticas_Basicas <- data.frame(RMS, Media, Mediana,
Desvio_Padrao)
  # Saída do resultado
  return(saida_Estatisticas_Basicas)
}

```

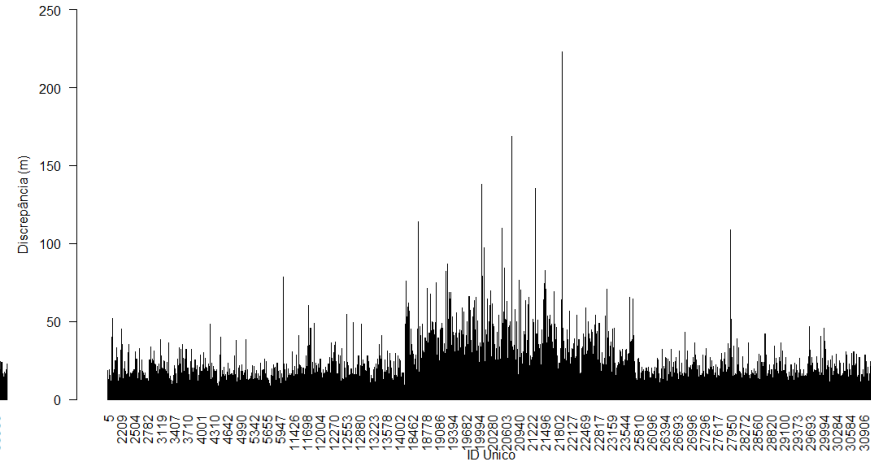
### Apêndice H. Gráficos das Discrepâncias das Populações



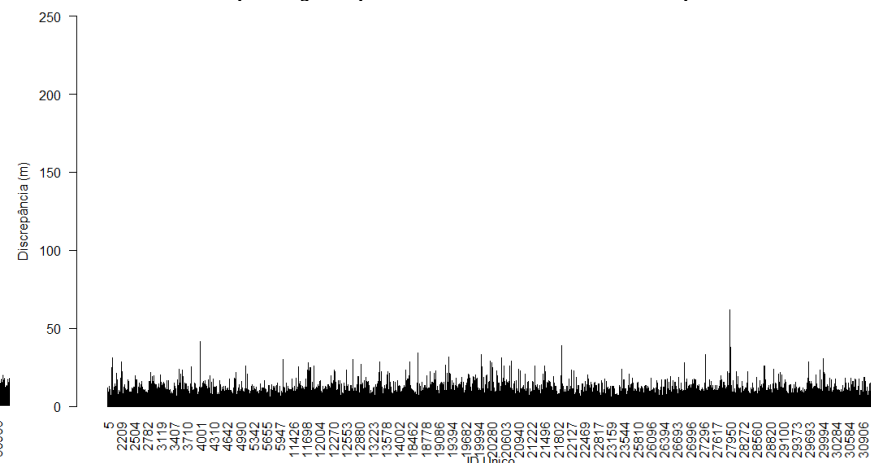
População para 1 km<sup>2</sup> – Buffer Simples



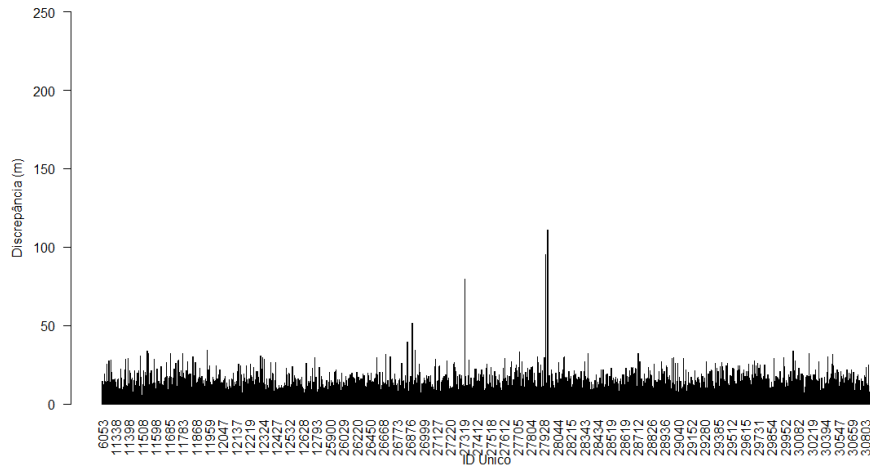
População para 1 km<sup>2</sup> – Distância de Hausdorff



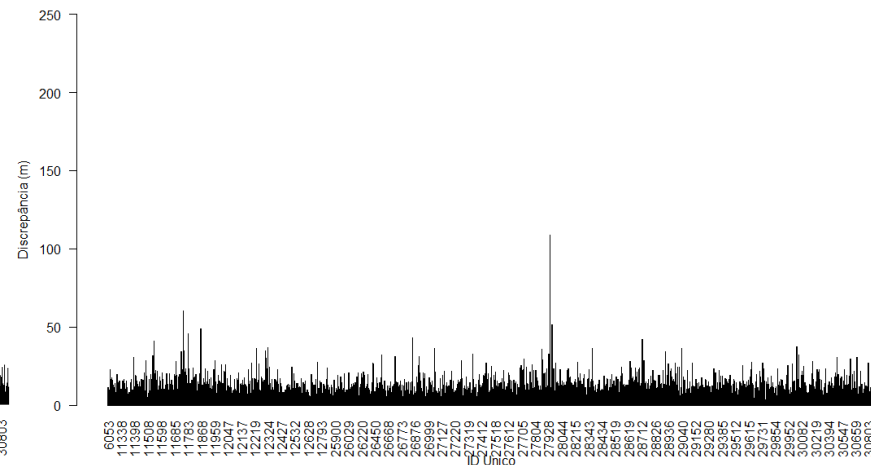
População para 1 km<sup>2</sup> – Buffer Duplo



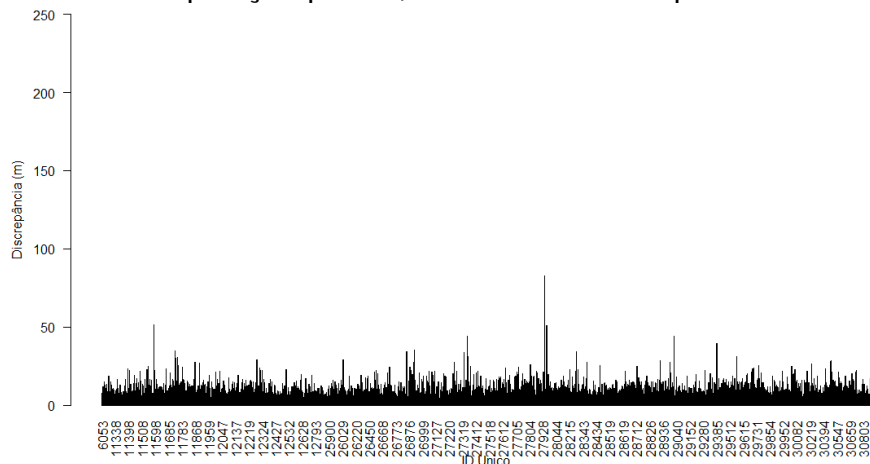
População para 1 km<sup>2</sup> – Influência do Vértice



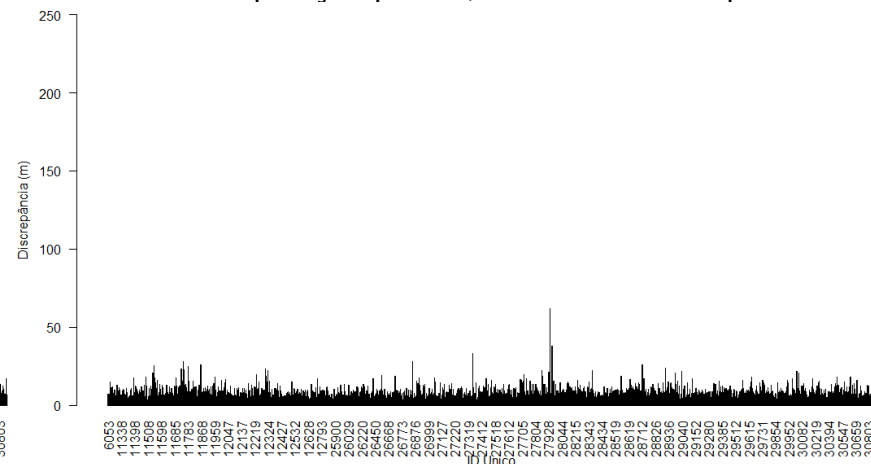
População para 2,5 km<sup>2</sup> – *Buffer Simples*



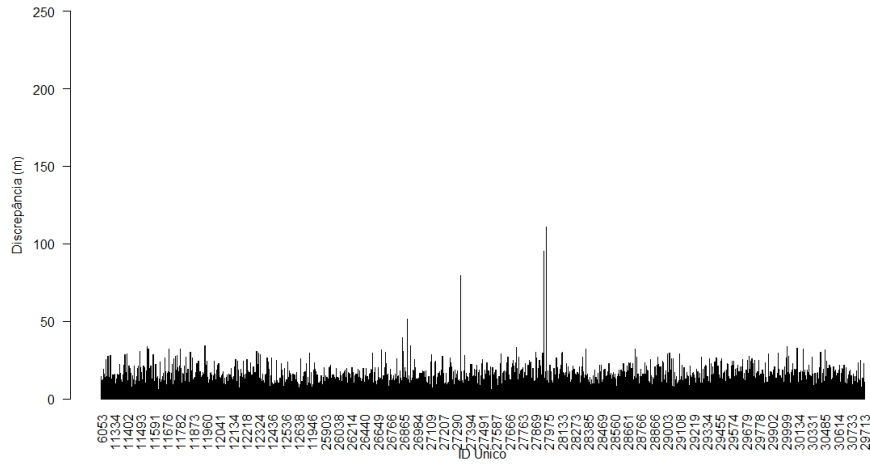
População para 2,5 km<sup>2</sup> – *Buffer Duplo*



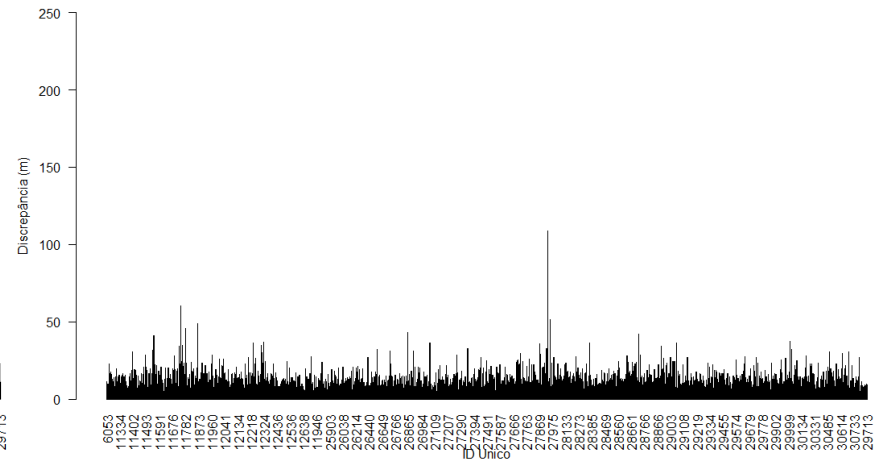
População para 2,5 km<sup>2</sup> – Distância de Hausdorff



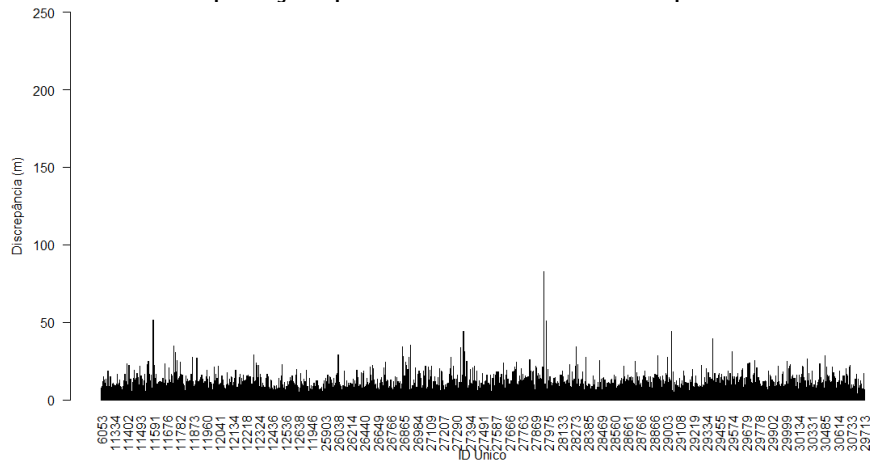
População para 2,5 km<sup>2</sup> – Influência do Vértice



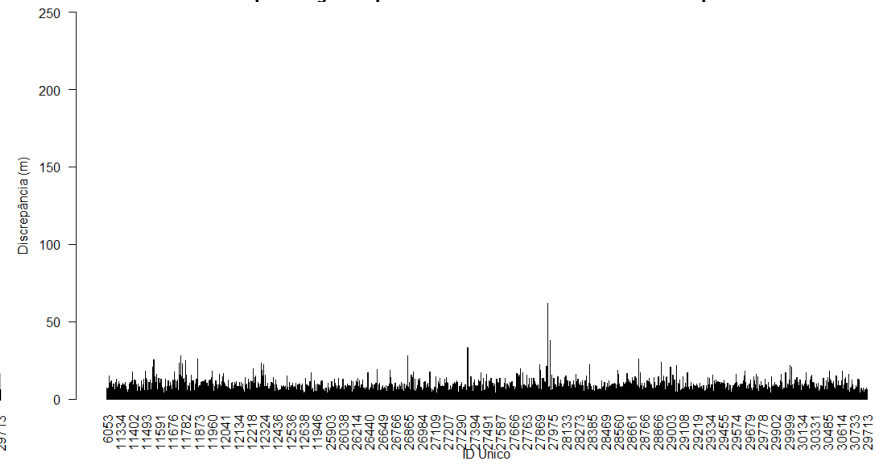
População para 5 km<sup>2</sup> – *Buffer Simples*



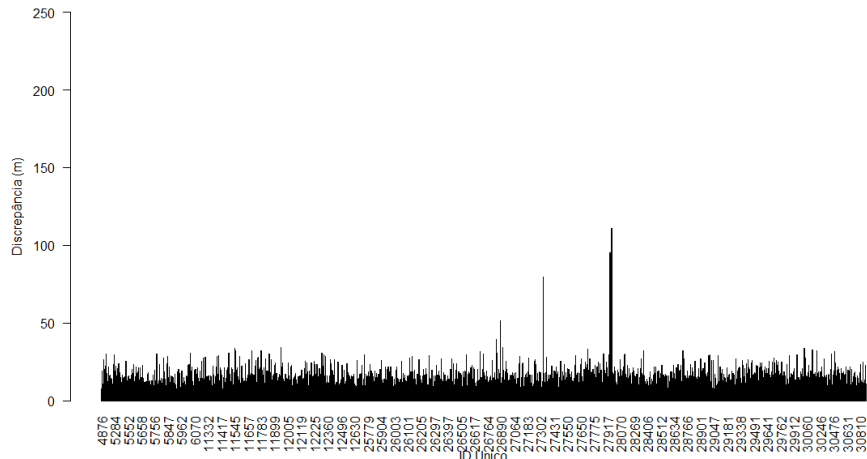
População para 5 km<sup>2</sup> – *Buffer Duplo*



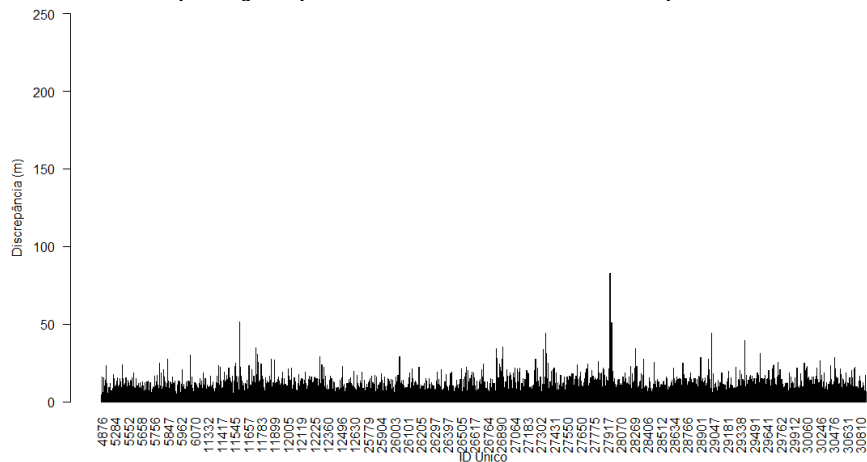
População para 5 km<sup>2</sup> – *Distância de Hausdorff*



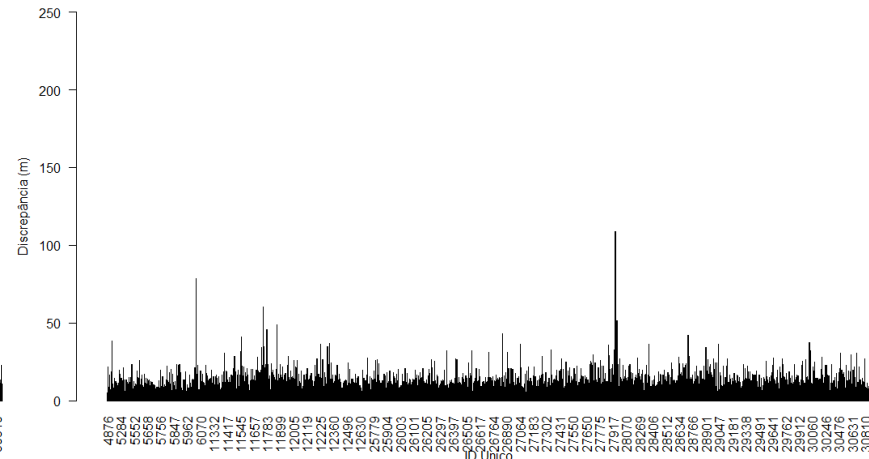
População para 5 km<sup>2</sup> – *Influência do Vértice*



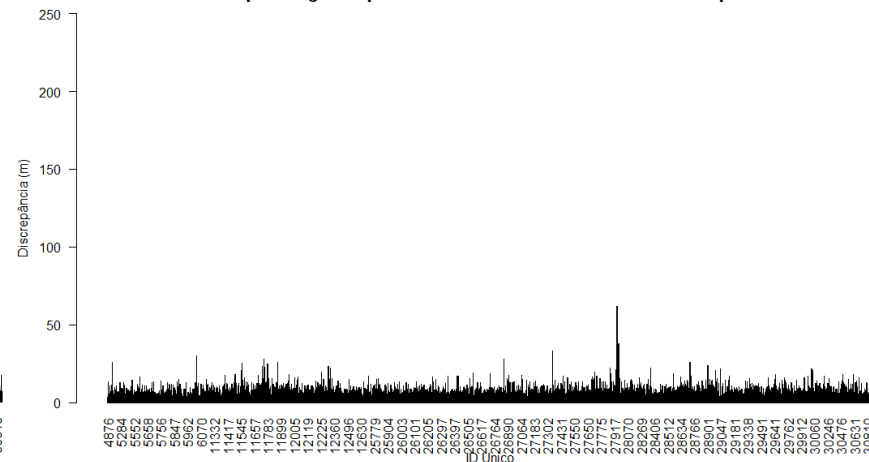
População para 7,5 km<sup>2</sup> – Buffer Simples



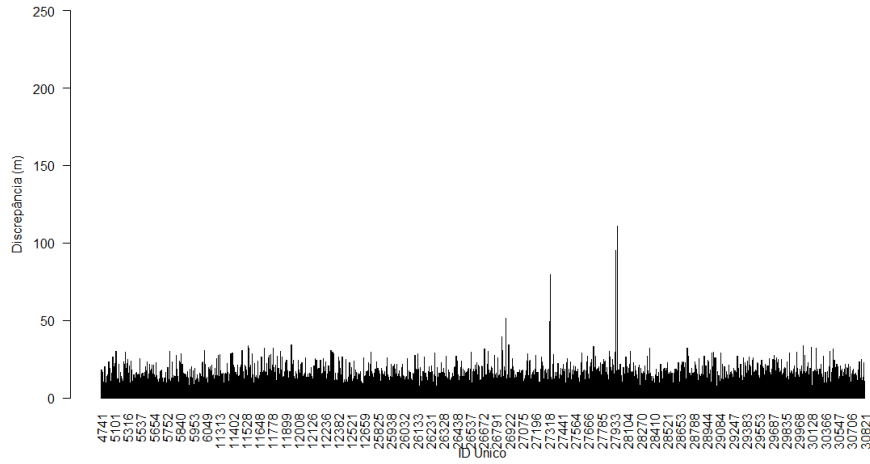
População para 7,5 km<sup>2</sup> – Distância de Hausdorff



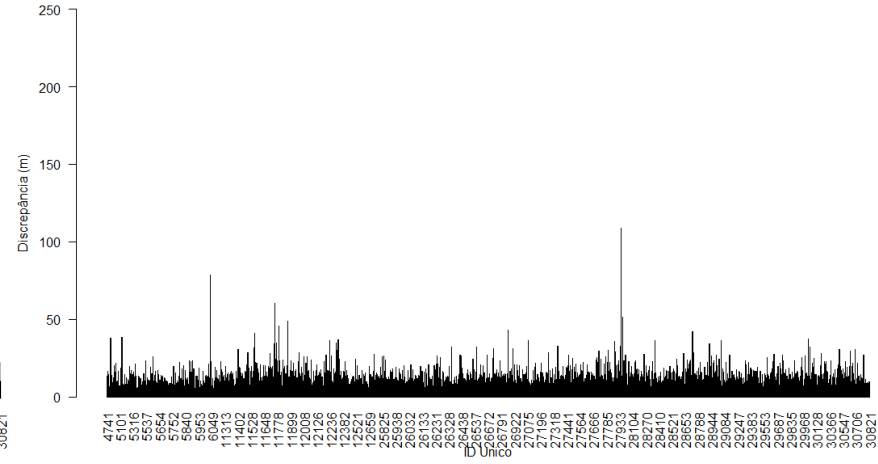
População para 7,5 km<sup>2</sup> – Buffer Duplo



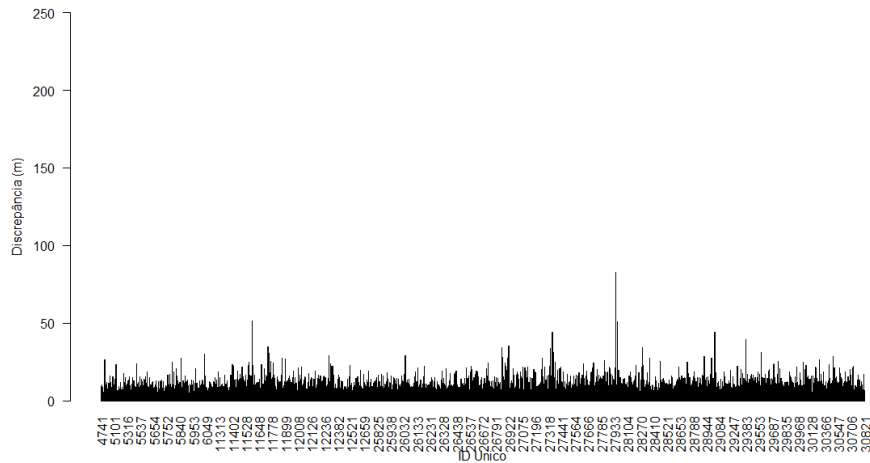
População para 7,5 km<sup>2</sup> – Influência do Vértice



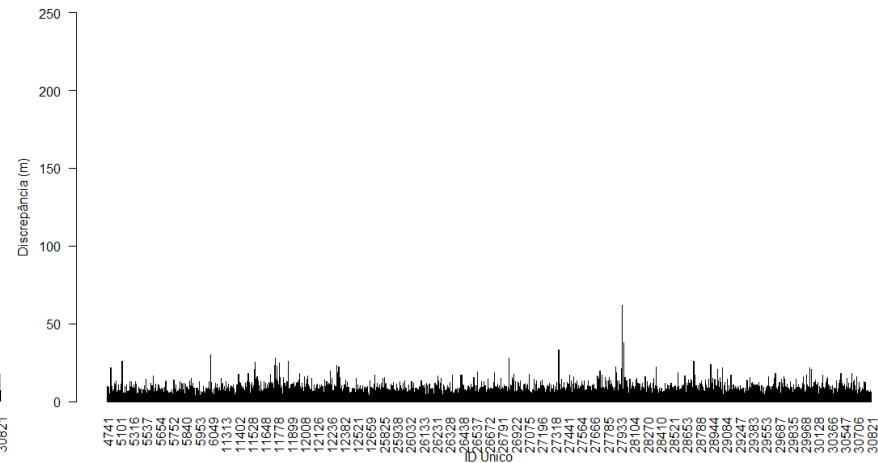
População para 10 km<sup>2</sup> – Buffer Simples



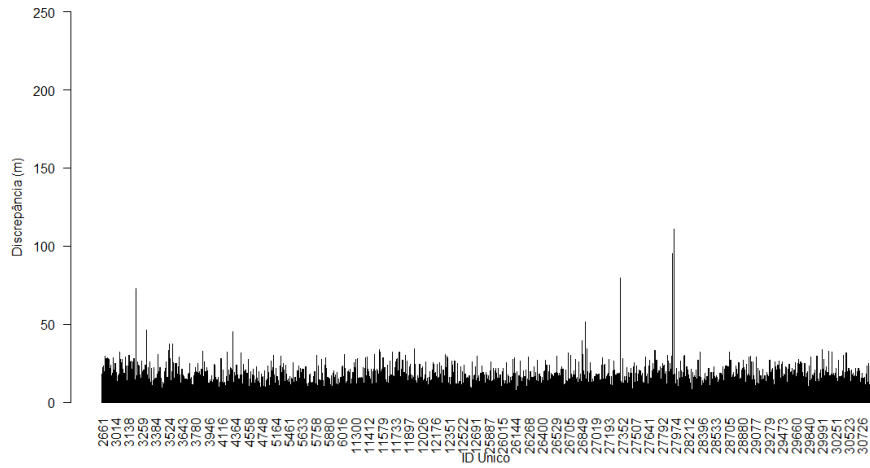
População para 10 km<sup>2</sup> – Buffer Duplo



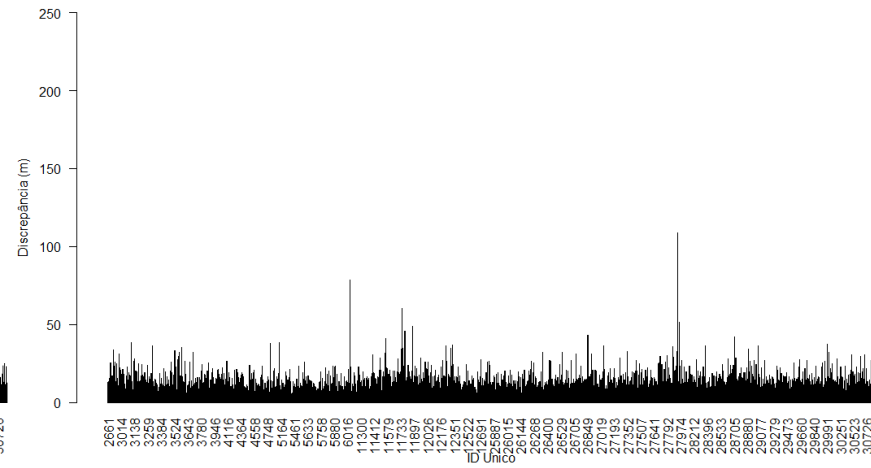
População para 10 km<sup>2</sup> – Distância de Hausdorff



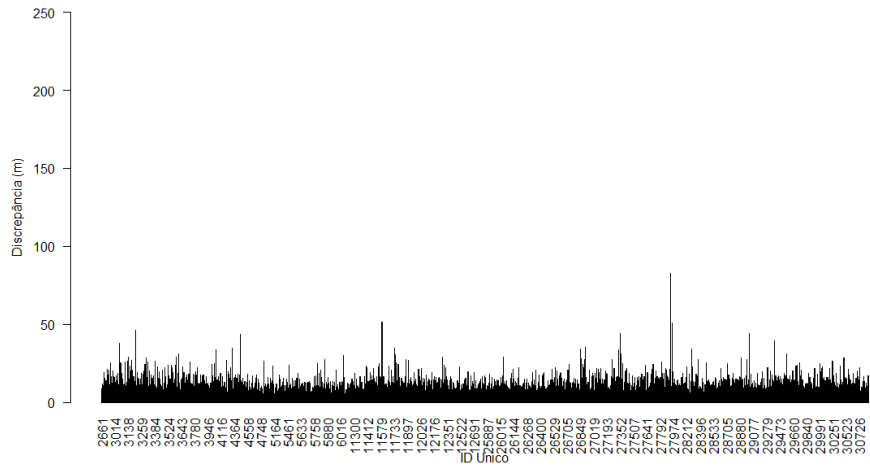
População para 10 km<sup>2</sup> – Influência do Vértice



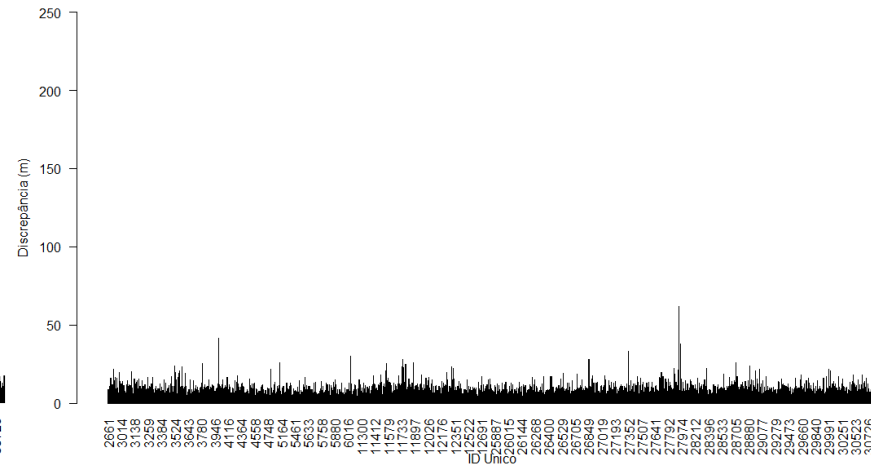
População para 15 km<sup>2</sup> – Buffer Simples



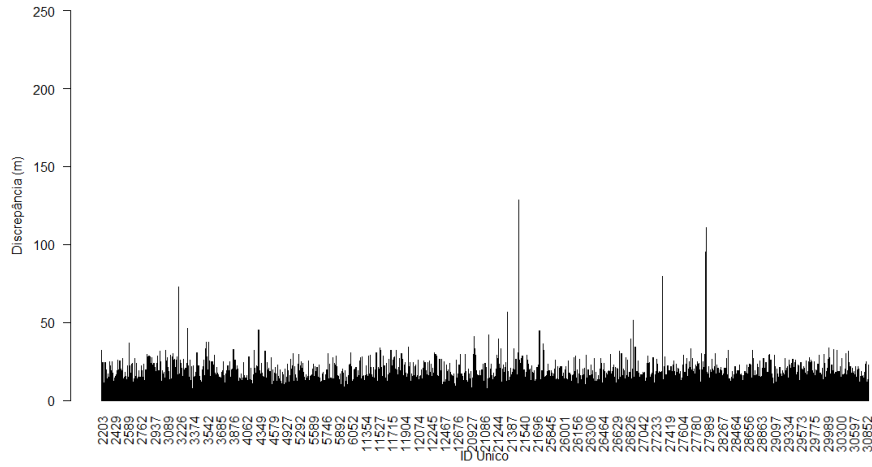
População para 15 km<sup>2</sup> – Buffer Duplo



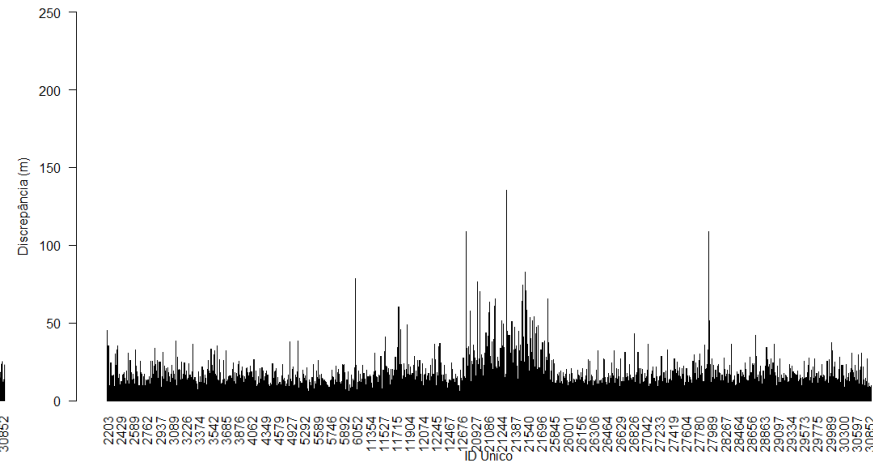
População para 15 km<sup>2</sup> – Distância de Hausdorff



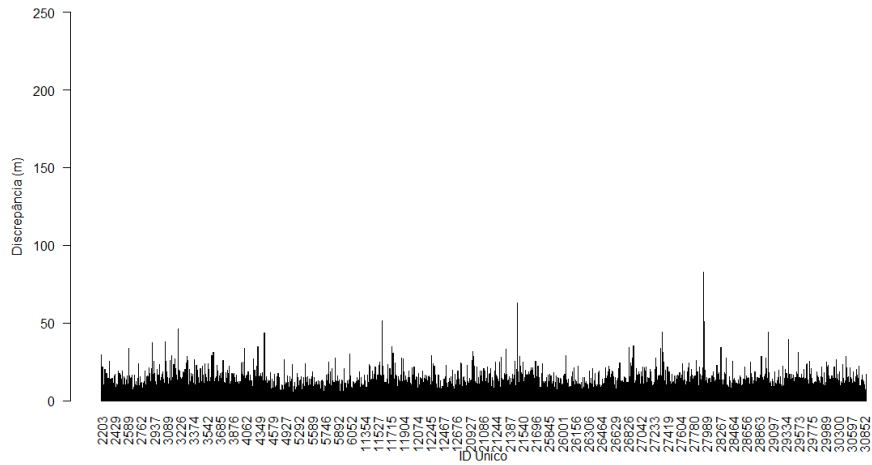
População para 15 km<sup>2</sup> – Influência do Vértice



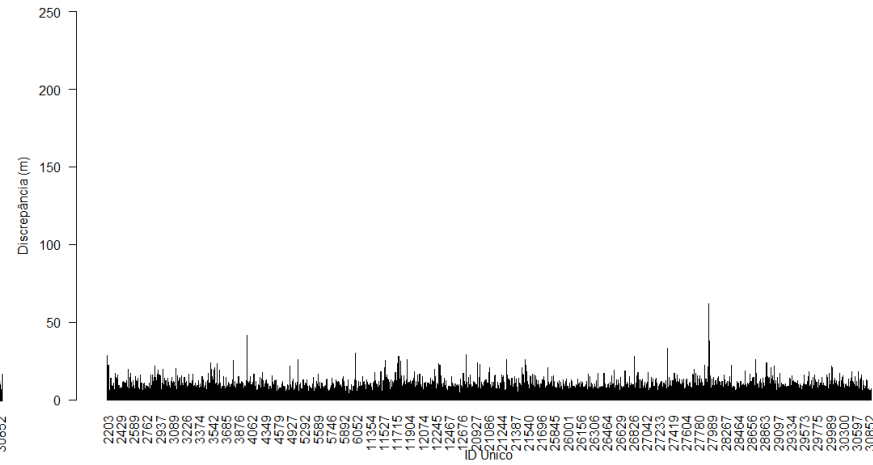
População para 20 km<sup>2</sup> – *Buffer Simples*



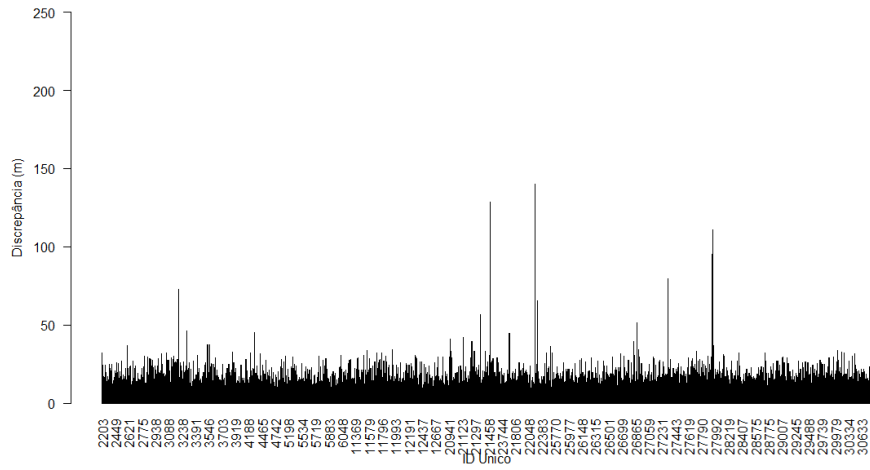
População para 20 km<sup>2</sup> – *Buffer Duplo*



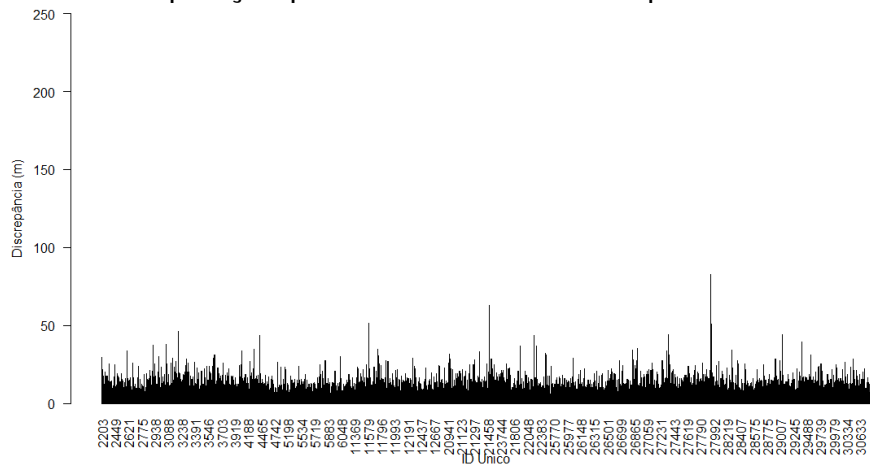
População para 20 km<sup>2</sup> – *Distância de Hausdorff*



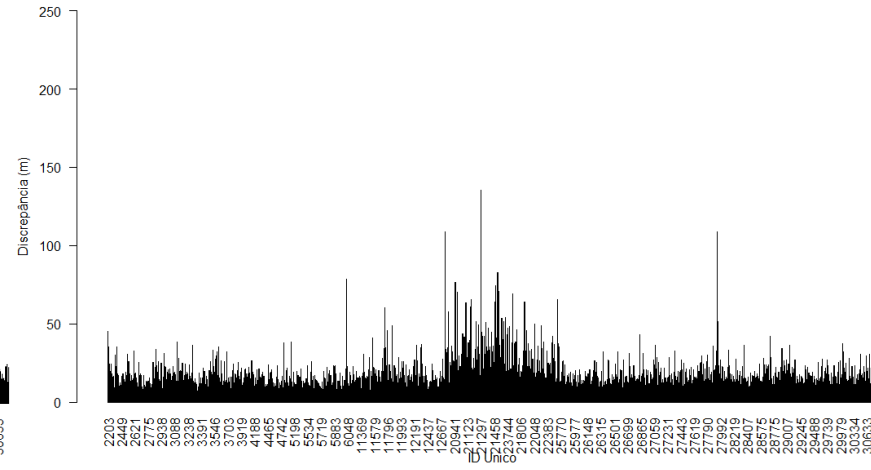
População para 20 km<sup>2</sup> – *Influência do Vértice*



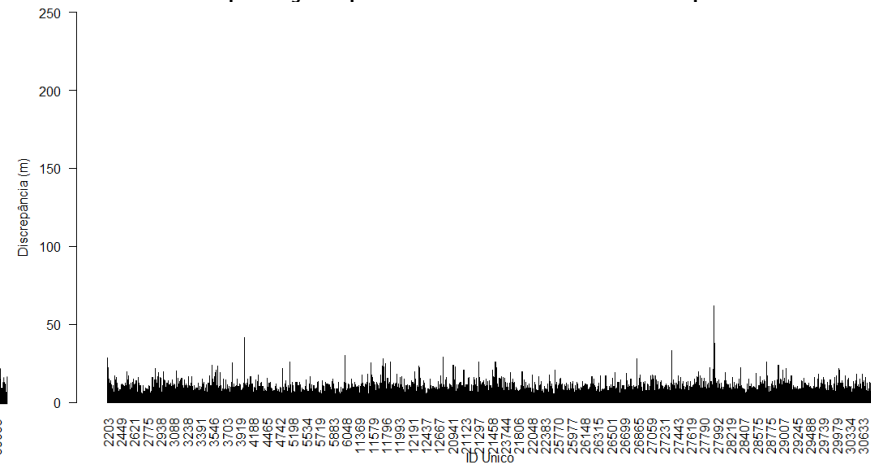
População para 25 km<sup>2</sup> – *Buffer Simples*



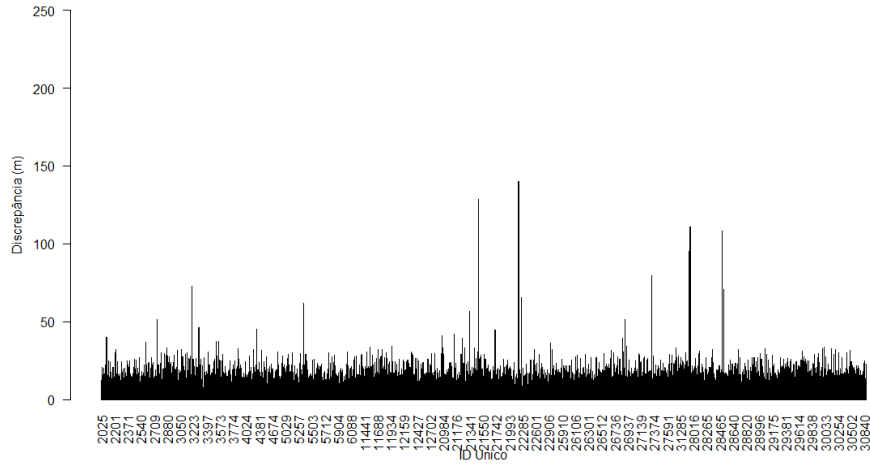
População para 25 km<sup>2</sup> – Distância de Hausdorff



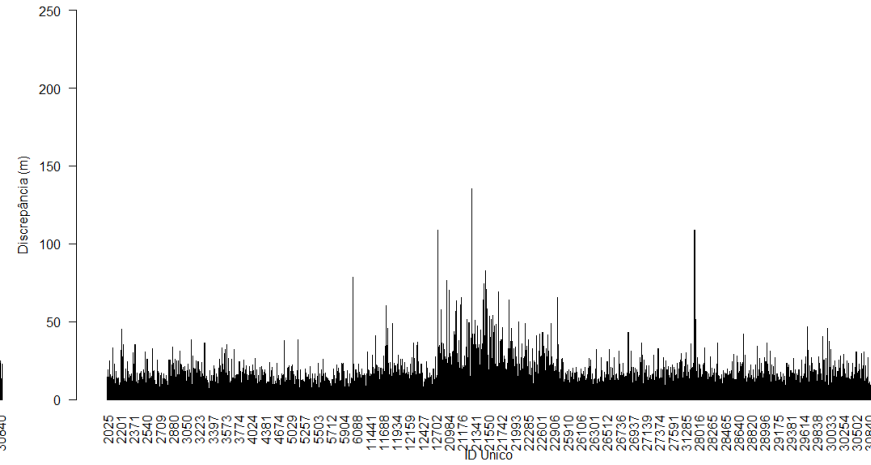
População para 25 km<sup>2</sup> – *Buffer Duplo*



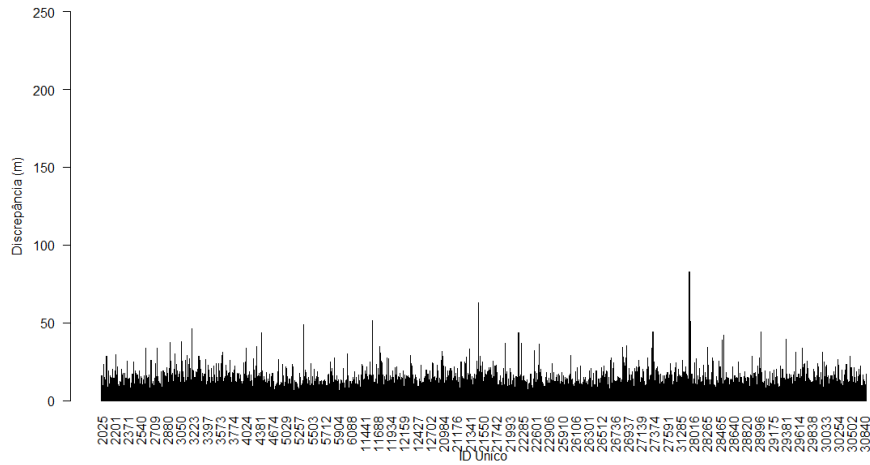
População para 25 km<sup>2</sup> – Influência do Vértice



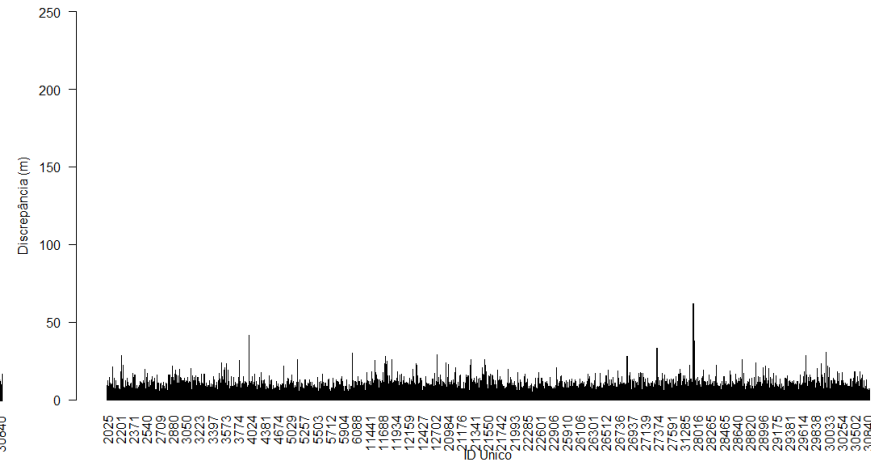
População para 50 km<sup>2</sup> – Buffer Simples



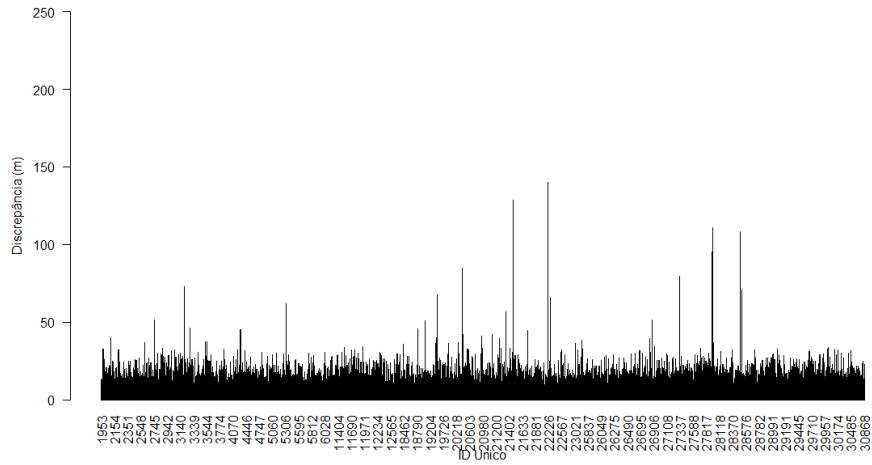
População para 50 km<sup>2</sup> – Buffer Duplo



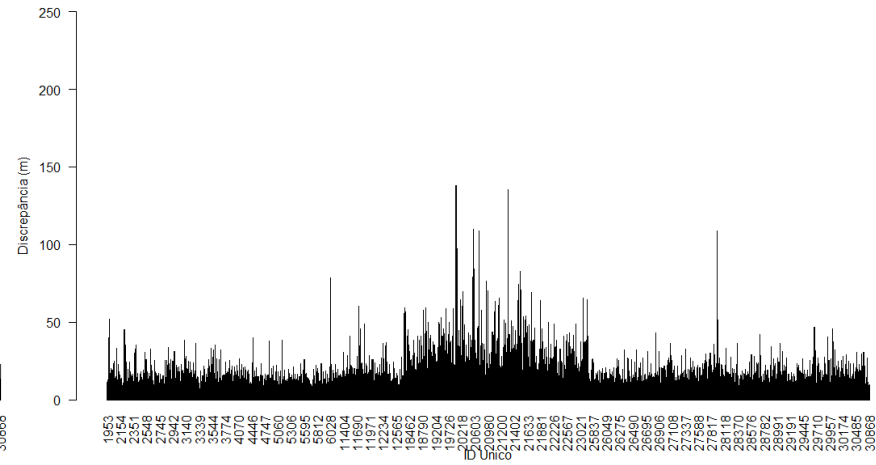
População para 50 km<sup>2</sup> – Distância de Hausdorff



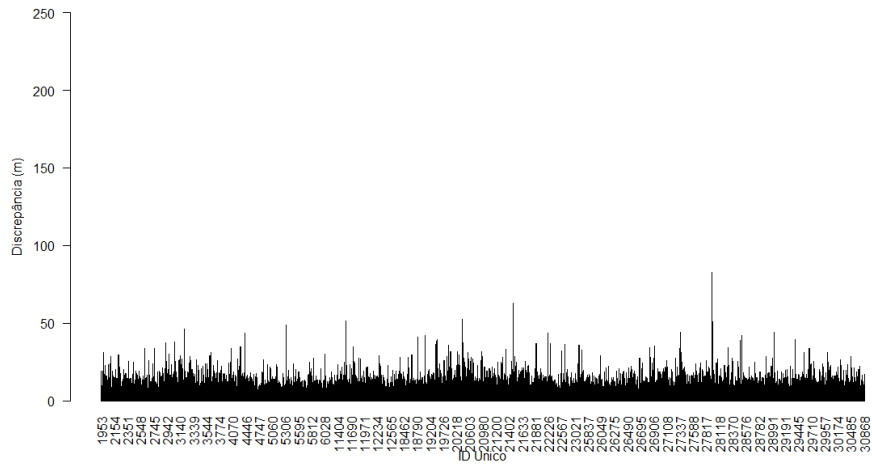
População para 50 km<sup>2</sup> – Influência do Vértice



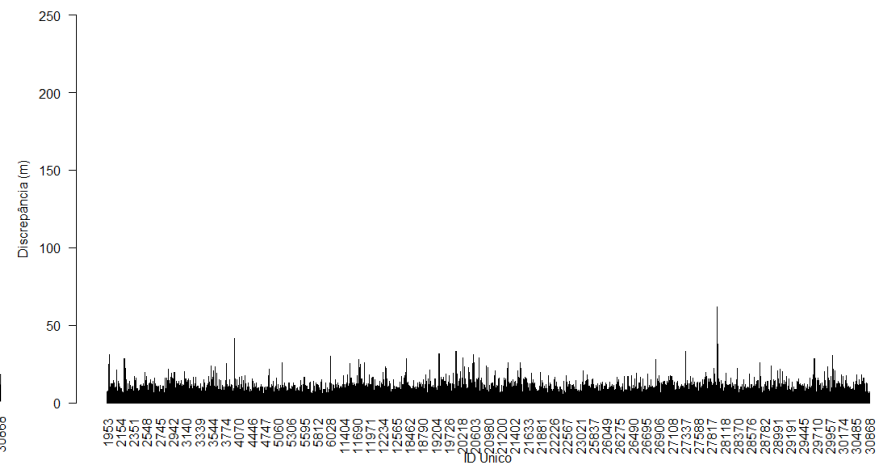
População para 75 km<sup>2</sup> – Buffer Simples



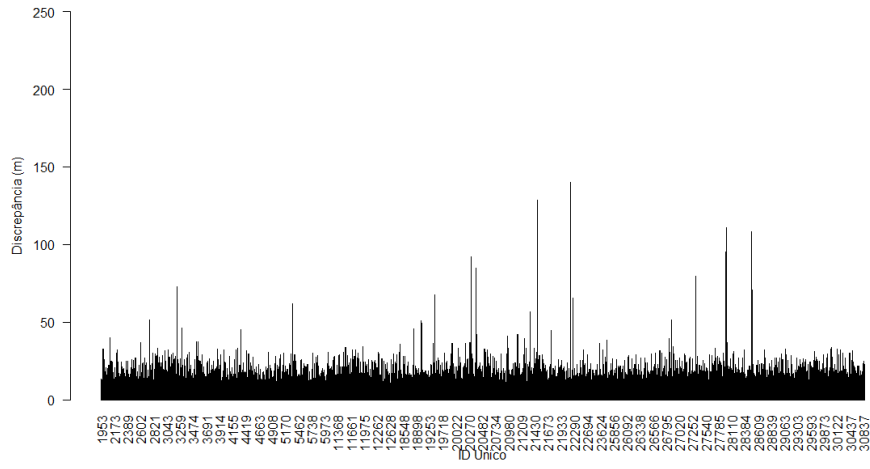
População para 75 km<sup>2</sup> – Buffer Duplo



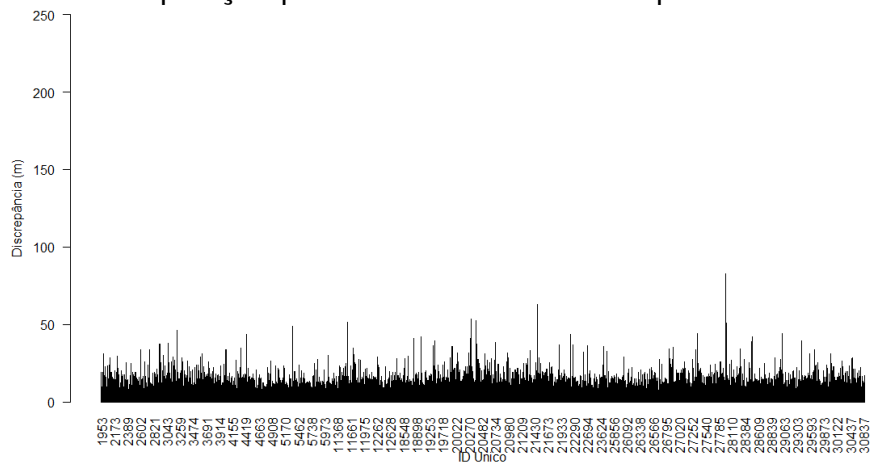
População para 75 km<sup>2</sup> – Distância de Hausdorff



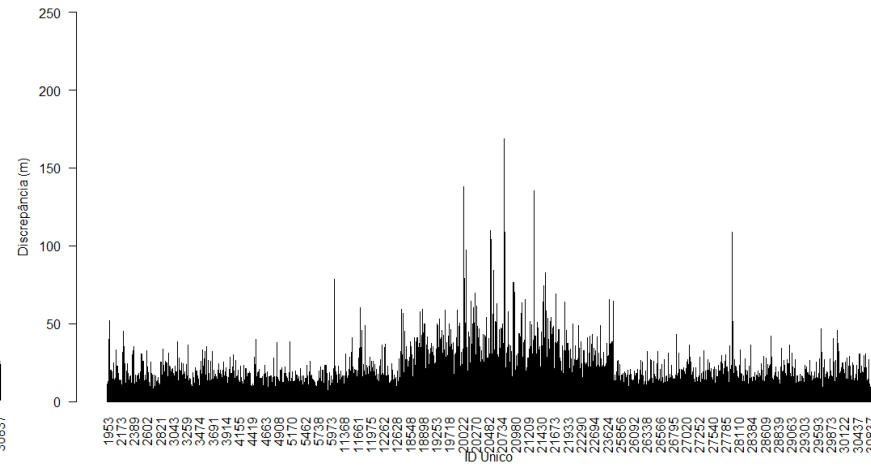
População para 75 km<sup>2</sup> – Influência do Vértice



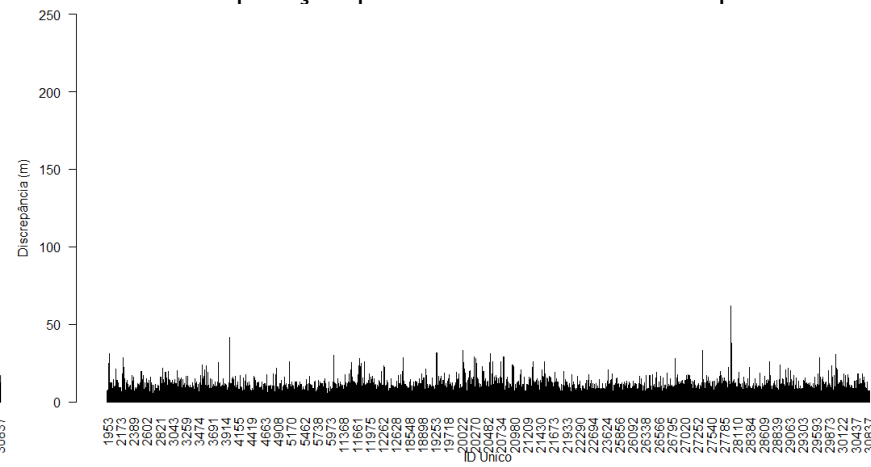
População para 100 km<sup>2</sup> – *Buffer Simples*



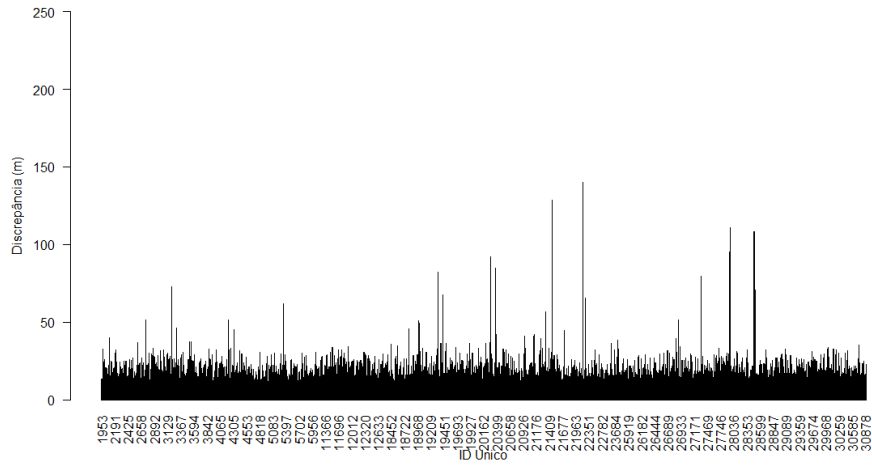
População para 100 km<sup>2</sup> – Distância de Hausdorff



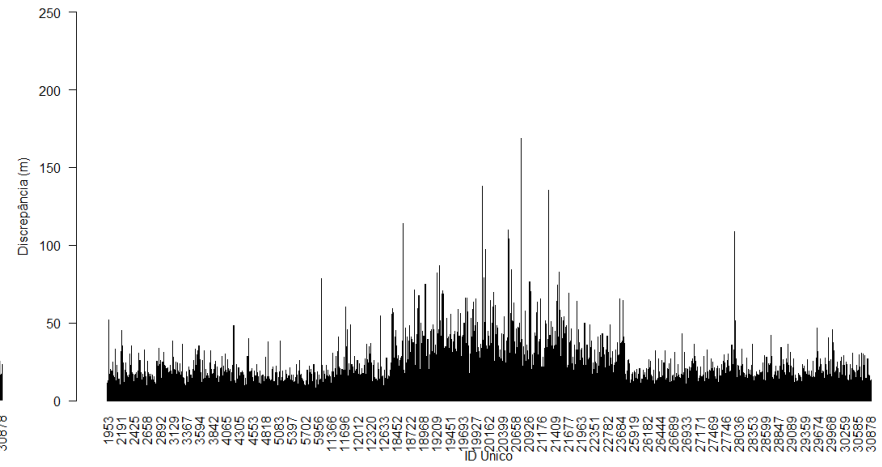
População para 100 km<sup>2</sup> – *Buffer Duplo*



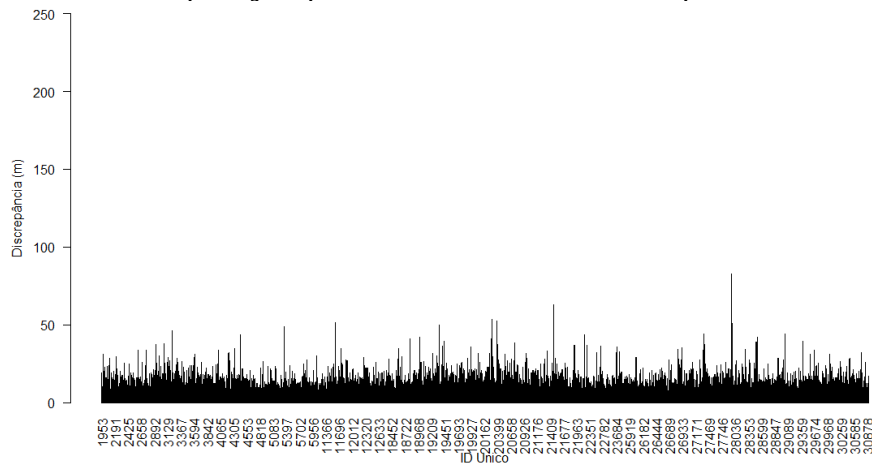
População para 100 km<sup>2</sup> – Influência do Vértice



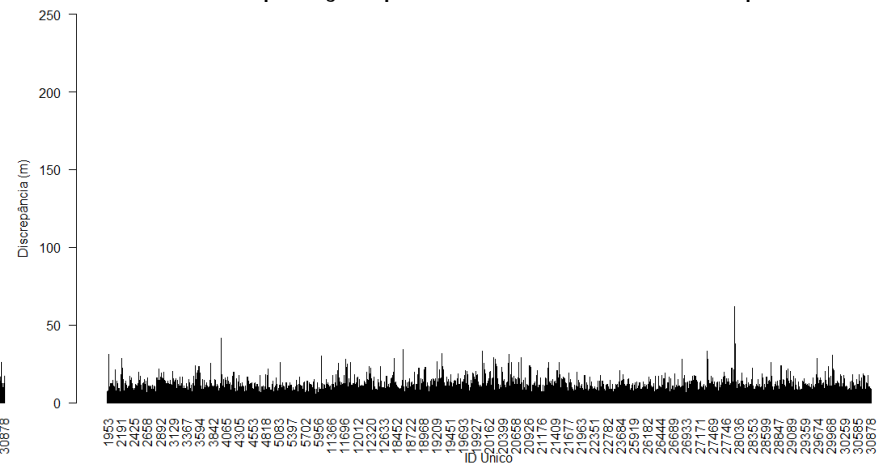
População para 125 km<sup>2</sup> – *Buffer Simples*



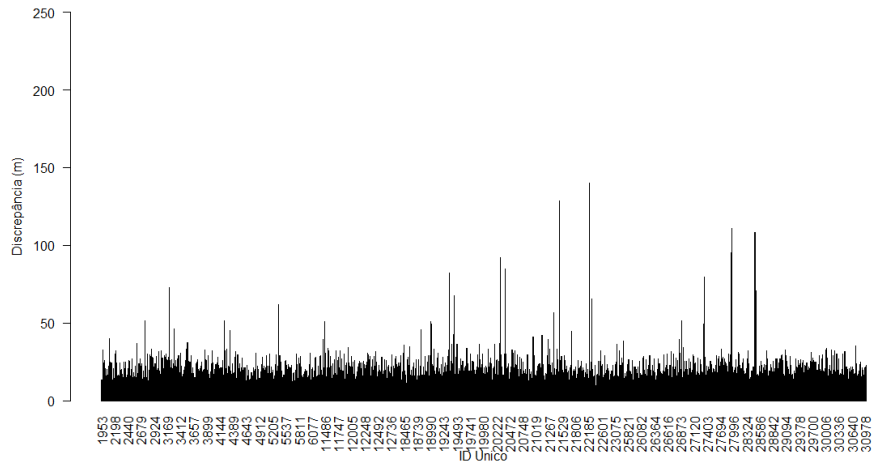
População para 125 km<sup>2</sup> – *Buffer Duplo*



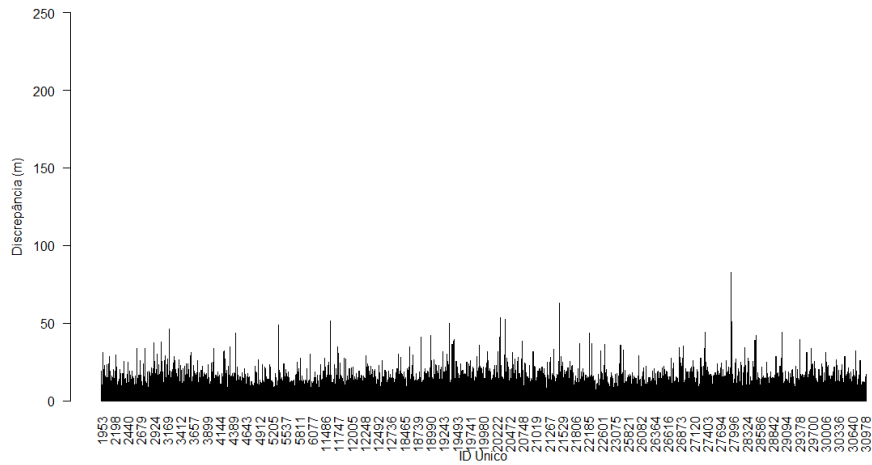
População para 125 km<sup>2</sup> – *Distância de Hausdorff*



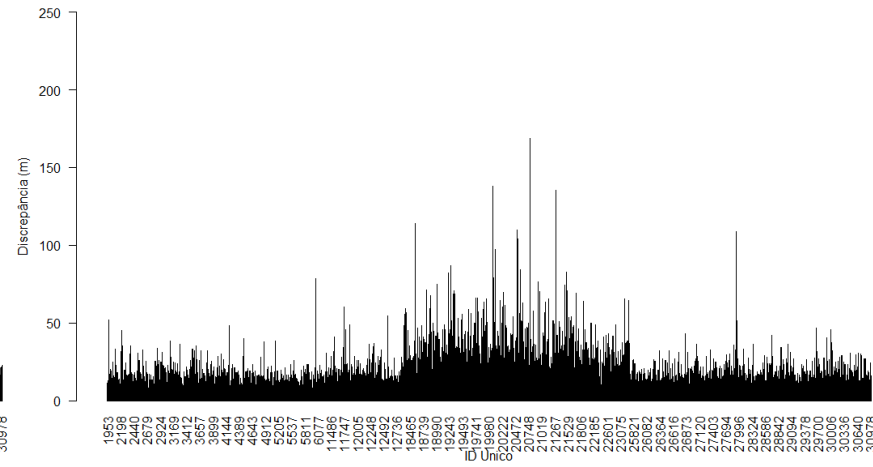
População para 125 km<sup>2</sup> – *Influência do Vértice*



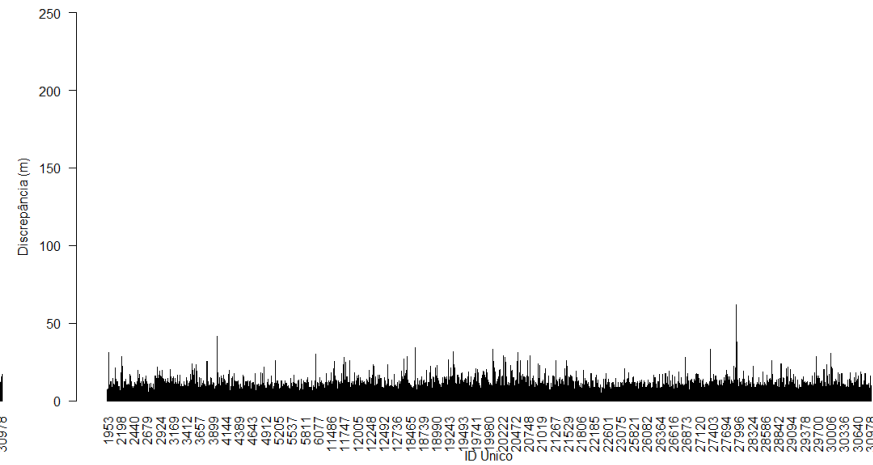
População para 150 km<sup>2</sup> – *Buffer Simples*



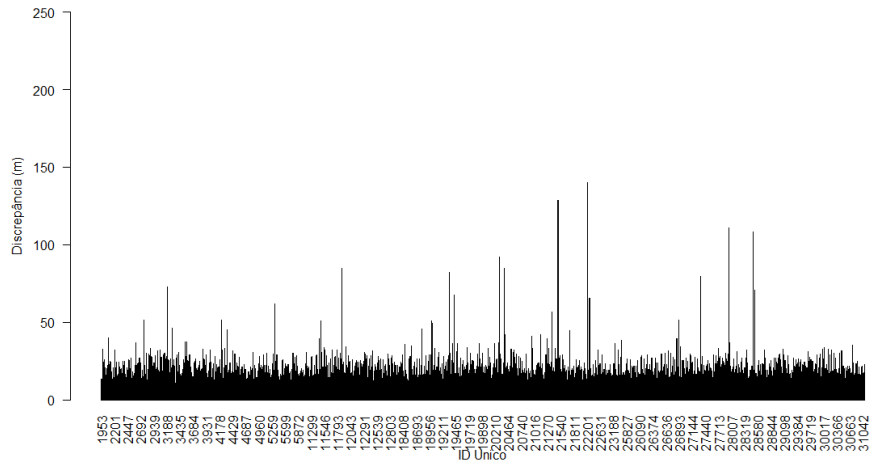
População para 150 km<sup>2</sup> – Distância de Hausdorff



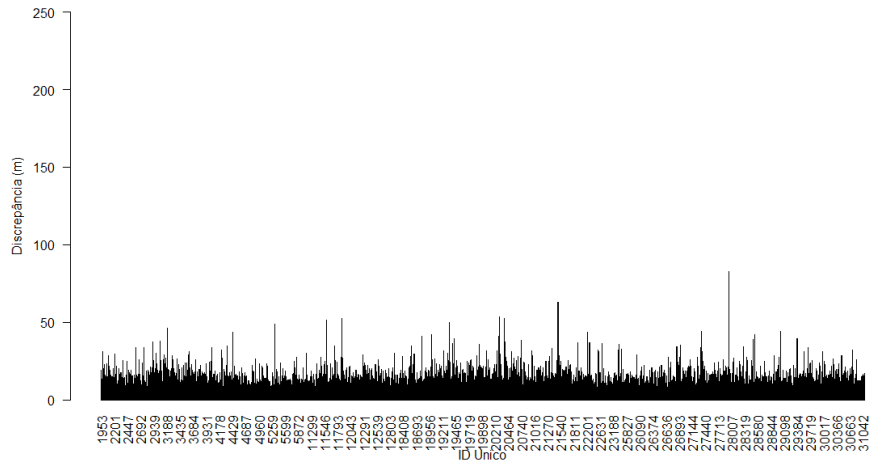
População para 150 km<sup>2</sup> – *Buffer Duplo*



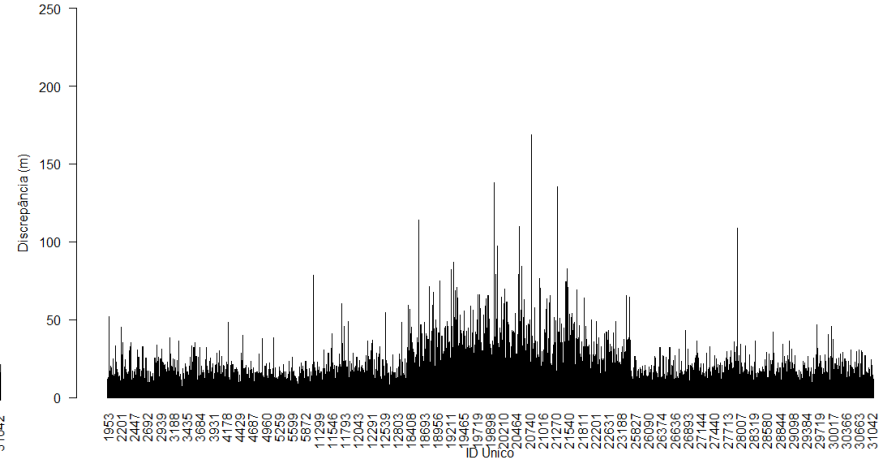
População para 150 km<sup>2</sup> – Influência do Vértice



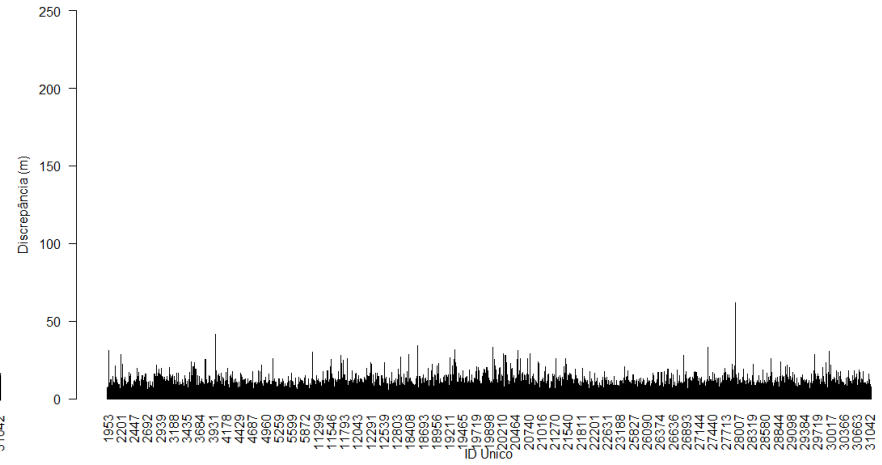
População para 175 km<sup>2</sup> – *Buffer Simples*



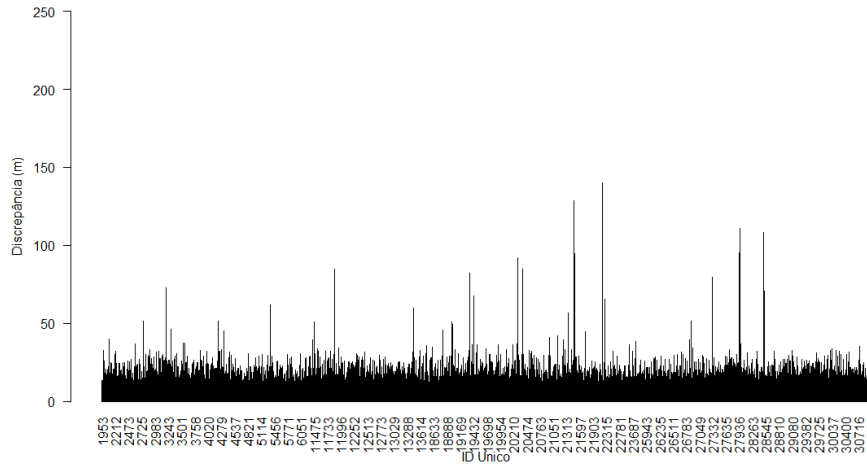
População para 175 km<sup>2</sup> – Distância de Hausdorff



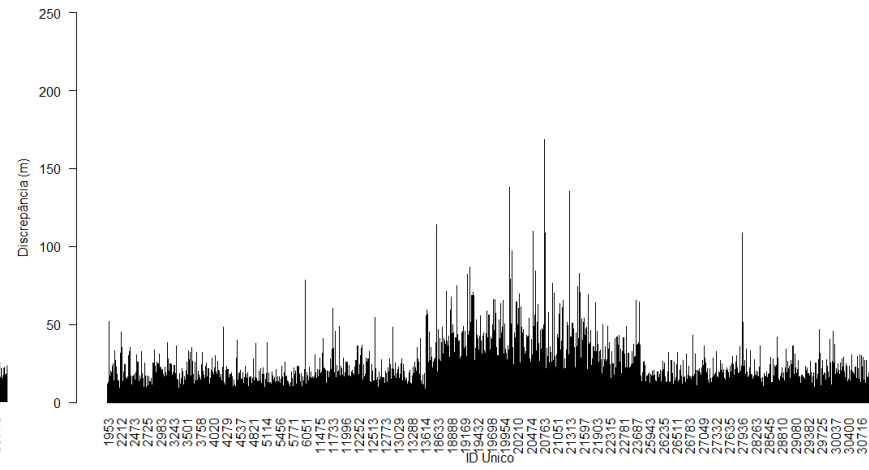
População para 175 km<sup>2</sup> – *Buffer Duplo*



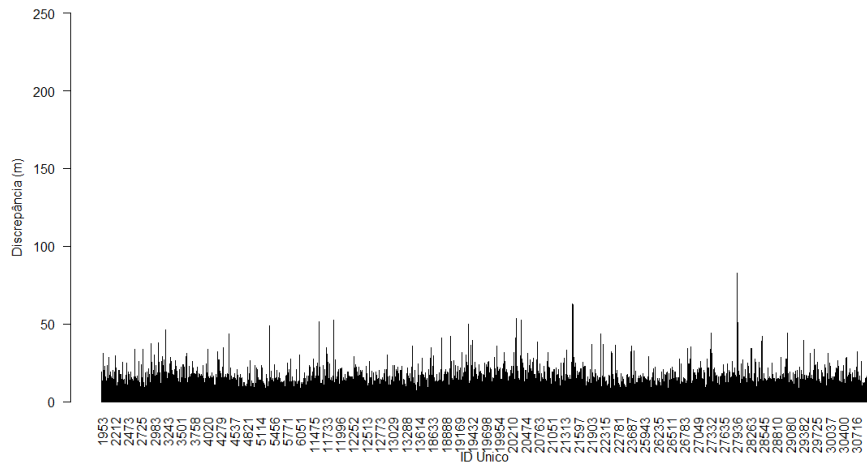
População para 175 km<sup>2</sup> – Influência do Vértice



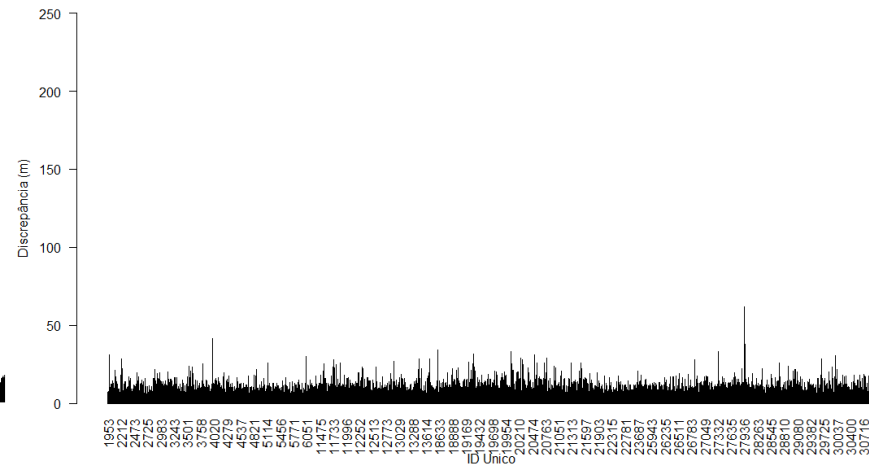
População para 200 km<sup>2</sup> – *Buffer Simples*



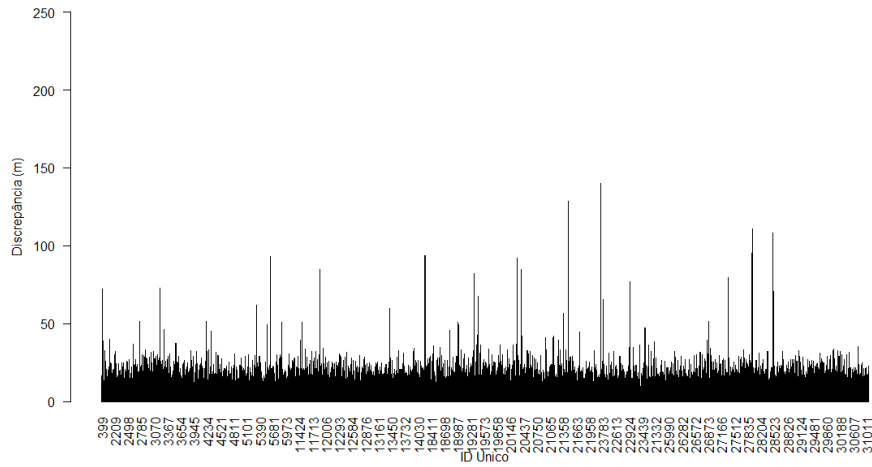
População para 200 km<sup>2</sup> – *Buffer Duplo*



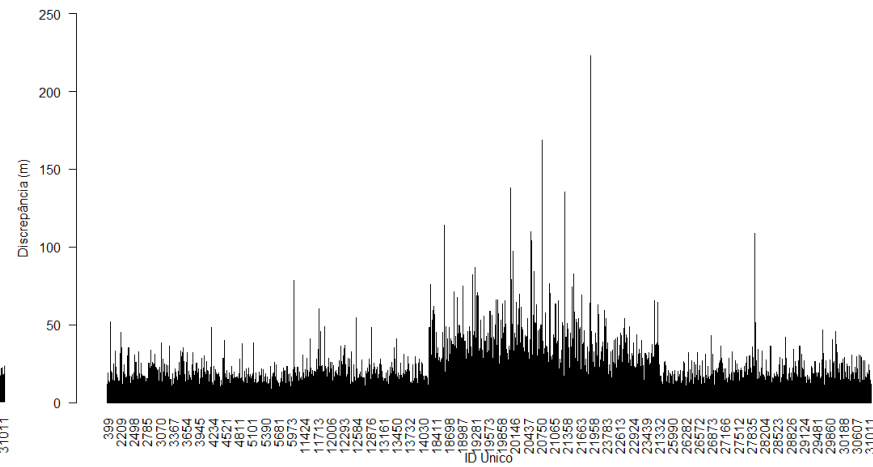
População para 200 km<sup>2</sup> – Distância de Hausdorff



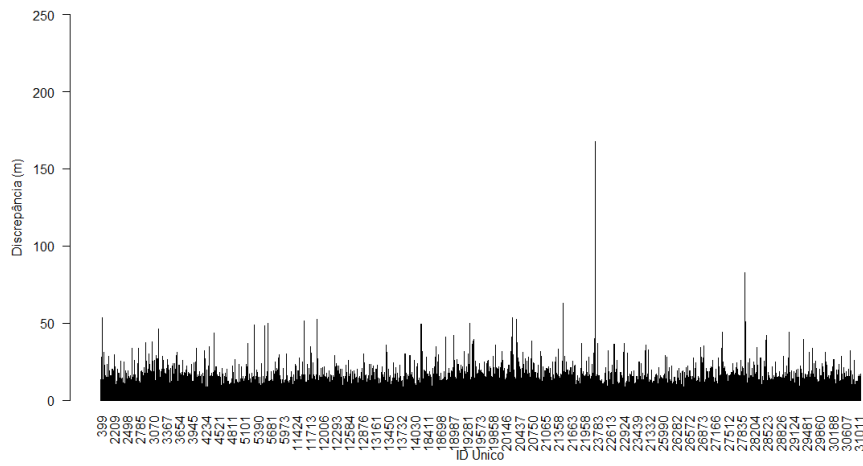
População para 200 km<sup>2</sup> – Influência do Vértice



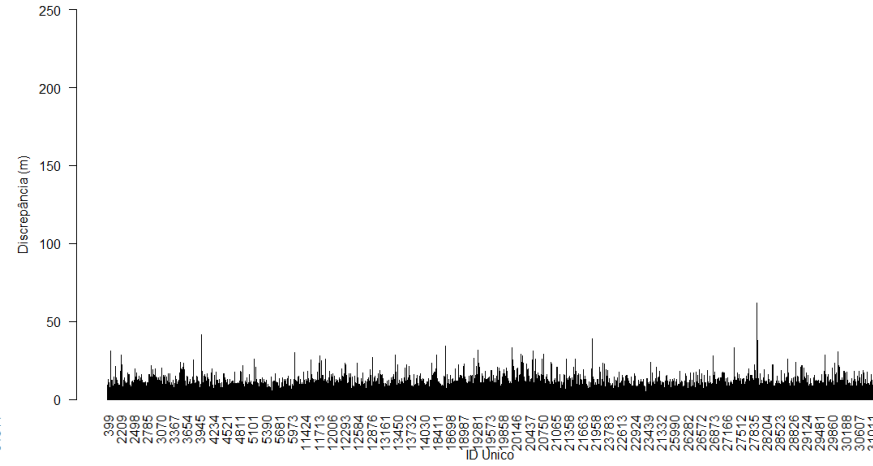
População para 225 km<sup>2</sup> – Buffer Simples



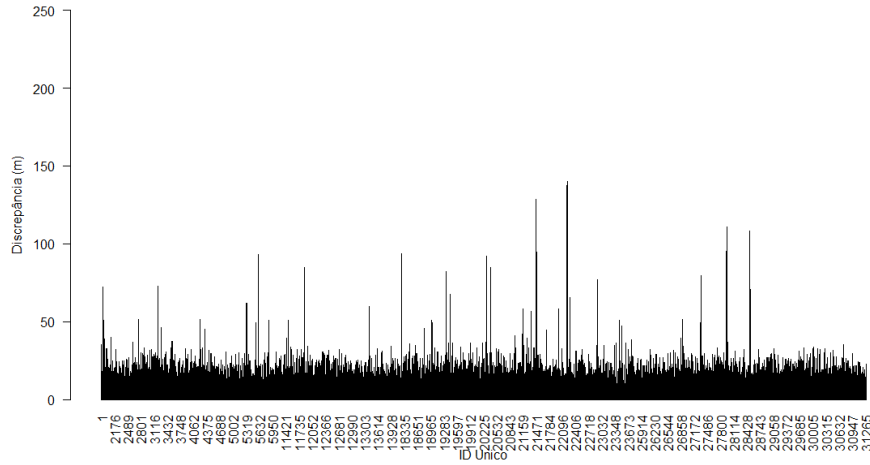
População para 225 km<sup>2</sup> – Buffer Duplo



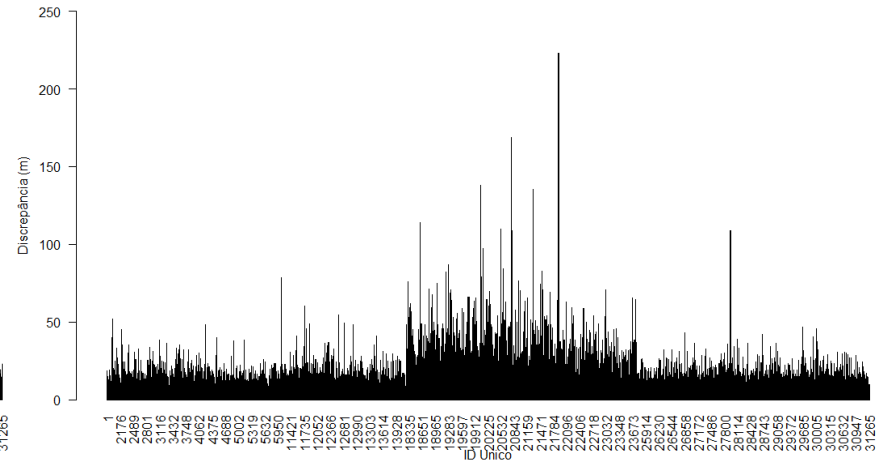
População para 225 km<sup>2</sup> – Distância de Hausdorff



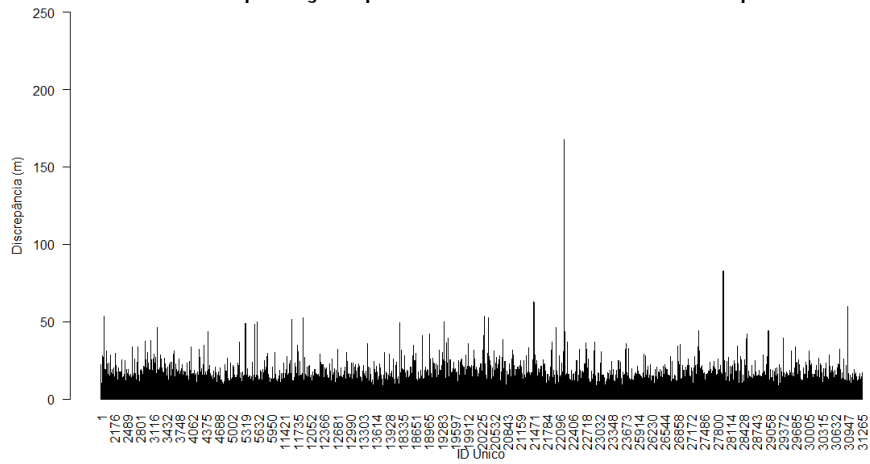
População para 225 km<sup>2</sup> – Influência do Vértice



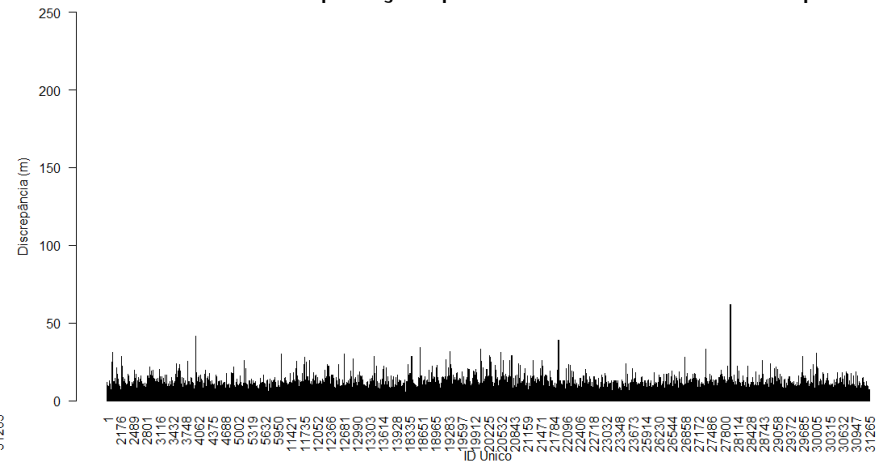
População para 250 km<sup>2</sup> – Buffer Simples



População para 250 km<sup>2</sup> – Buffer Duplo



População para 250 km<sup>2</sup> – Distância de Hausdorff



População para 250 km<sup>2</sup> – Influência do Vértice

## Apêndice I. Estatísticas das Discrepâncias das Populações

| População para 1 km <sup>2</sup> |         |         |        |        |
|----------------------------------|---------|---------|--------|--------|
| Estatística (m)                  | BD      | BS      | DH     | IV     |
| Mínimo                           | 0,016   | 0,268   | 0,183  | 0,118  |
| Máximo                           | 223,006 | 140,076 | 82,728 | 61,765 |
| Média                            | 10,464  | 10,613  | 8,475  | 5,984  |
| Desvio-Padrão                    | 9,248   | 6,735   | 5,117  | 3,594  |
| RMS                              | 13,965  | 12,570  | 9,901  | 6,981  |

| População para 2,5 km <sup>2</sup> |         |         |        |        |
|------------------------------------|---------|---------|--------|--------|
| Estatística (m)                    | BD      | BS      | DH     | IV     |
| Mínimo                             | 0,002   | 0,384   | 0,751  | 0,464  |
| Máximo                             | 108,618 | 110,784 | 82,728 | 61,765 |
| Média                              | 9,069   | 10,813  | 8,468  | 6,132  |
| Desvio-Padrão                      | 5,884   | 6,171   | 4,856  | 3,584  |
| RMS                                | 10,810  | 12,450  | 9,761  | 7,103  |

| População para 5 km <sup>2</sup> |         |         |        |        |
|----------------------------------|---------|---------|--------|--------|
| Estatística (m)                  | BD      | BS      | DH     | IV     |
| Mínimo                           | 0,002   | 0,384   | 0,751  | 0,464  |
| Máximo                           | 108,618 | 110,784 | 82,728 | 61,765 |
| Média                            | 9,047   | 10,869  | 8,414  | 6,128  |
| Desvio-Padrão                    | 5,787   | 6,096   | 4,769  | 3,526  |
| RMS                              | 10,740  | 12,462  | 9,671  | 7,070  |

| População para 7,5 km <sup>2</sup> |         |         |        |        |
|------------------------------------|---------|---------|--------|--------|
| Estatística (m)                    | BD      | BS      | DH     | IV     |
| Mínimo                             | 0,002   | 0,384   | 0,597  | 0,250  |
| Máximo                             | 108,618 | 110,784 | 82,728 | 61,765 |
| Média                              | 8,804   | 10,720  | 8,194  | 5,985  |
| Desvio-Padrão                      | 5,663   | 5,934   | 4,610  | 3,421  |
| RMS                                | 10,468  | 12,253  | 9,402  | 6,894  |

| População para 10 km <sup>2</sup> |         |         |        |        |
|-----------------------------------|---------|---------|--------|--------|
| Estatística (m)                   | BD      | BS      | DH     | IV     |
| Mínimo                            | 0,002   | 0,384   | 0,597  | 0,250  |
| Máximo                            | 108,618 | 110,784 | 82,728 | 61,765 |
| Média                             | 8,750   | 10,703  | 8,095  | 5,954  |
| Desvio-Padrão                     | 5,609   | 5,855   | 4,559  | 3,382  |
| RMS                               | 10,394  | 12,200  | 9,290  | 6,847  |

População para 15 km<sup>2</sup>

| Estatística (m) | BD      | BS      | DH     | IV     |
|-----------------|---------|---------|--------|--------|
| Mínimo          | 0,002   | 0,384   | 0,597  | 0,250  |
| Máximo          | 108,618 | 110,784 | 82,728 | 61,765 |
| Média           | 8,763   | 10,796  | 8,212  | 5,987  |
| Desvio-Padrão   | 5,527   | 5,890   | 4,668  | 3,376  |
| RMS             | 10,360  | 12,298  | 9,447  | 6,873  |

População para 20 km<sup>2</sup>

| Estatística (m) | BD      | BS      | DH     | IV     |
|-----------------|---------|---------|--------|--------|
| Mínimo          | 0,002   | 0,384   | 0,597  | 0,250  |
| Máximo          | 135,394 | 128,813 | 82,728 | 61,765 |
| Média           | 9,301   | 10,703  | 8,220  | 5,955  |
| Desvio-Padrão   | 7,058   | 6,150   | 4,763  | 3,418  |
| RMS             | 11,676  | 12,344  | 9,501  | 6,866  |

População para 25 km<sup>2</sup>

| Estatística (m) | BD      | BS      | DH     | IV     |
|-----------------|---------|---------|--------|--------|
| Mínimo          | 0,002   | 0,384   | 0,597  | 0,250  |
| Máximo          | 135,394 | 140,076 | 82,728 | 61,765 |
| Média           | 9,349   | 10,661  | 8,162  | 5,895  |
| Desvio-Padrão   | 7,155   | 6,252   | 4,760  | 3,371  |
| RMS             | 11,773  | 12,359  | 9,449  | 6,791  |

População para 50 km<sup>2</sup>

| Estatística (m) | BD      | BS      | DH     | IV     |
|-----------------|---------|---------|--------|--------|
| Mínimo          | 0,002   | 0,384   | 0,409  | 0,250  |
| Máximo          | 135,394 | 140,076 | 82,728 | 61,765 |
| Média           | 9,225   | 10,564  | 8,058  | 5,814  |
| Desvio-Padrão   | 7,019   | 6,261   | 4,722  | 3,320  |
| RMS             | 11,592  | 12,280  | 9,340  | 6,695  |

População para 75 km<sup>2</sup>

| Estatística (m) | BD      | BS      | DH     | IV     |
|-----------------|---------|---------|--------|--------|
| Mínimo          | 0,002   | 0,268   | 0,190  | 0,118  |
| Máximo          | 137,779 | 140,076 | 82,728 | 61,765 |
| Média           | 9,828   | 10,593  | 8,162  | 5,888  |
| Desvio-Padrão   | 8,067   | 6,293   | 4,826  | 3,413  |
| RMS             | 12,715  | 12,321  | 9,482  | 6,806  |

População para 100 km<sup>2</sup>

| Estatística (m) | BD      | BS      | DH     | IV     |
|-----------------|---------|---------|--------|--------|
| Mínimo          | 0,002   | 0,233   | 0,190  | 0,118  |
| Máximo          | 169,000 | 140,076 | 82,728 | 61,765 |
| Média           | 10,053  | 10,575  | 8,137  | 5,890  |
| Desvio-Padrão   | 8,427   | 6,254   | 4,802  | 3,406  |
| RMS             | 13,118  | 12,286  | 9,449  | 6,804  |

População para 125 km<sup>2</sup>

| Estatística (m) | BD      | BS      | DH     | IV     |
|-----------------|---------|---------|--------|--------|
| Mínimo          | 0,002   | 0,233   | 0,190  | 0,118  |
| Máximo          | 169,000 | 140,076 | 82,728 | 61,765 |
| Média           | 10,374  | 10,671  | 8,206  | 5,945  |
| Desvio-Padrão   | 8,881   | 6,301   | 4,843  | 3,446  |
| RMS             | 13,657  | 12,393  | 9,529  | 6,872  |

População para 150 km<sup>2</sup>

| Estatística (m) | BD      | BS      | DH     | IV     |
|-----------------|---------|---------|--------|--------|
| Mínimo          | 0,002   | 0,233   | 0,190  | 0,118  |
| Máximo          | 169,000 | 140,076 | 82,728 | 61,765 |
| Média           | 10,291  | 10,669  | 8,161  | 5,930  |
| Desvio-Padrão   | 8,780   | 6,275   | 4,818  | 3,429  |
| RMS             | 13,528  | 12,378  | 9,477  | 6,850  |

População para 175 km<sup>2</sup>

| Estatística (m) | BD      | BS      | DH     | IV     |
|-----------------|---------|---------|--------|--------|
| Mínimo          | 0,002   | 0,233   | 0,190  | 0,118  |
| Máximo          | 169,000 | 140,076 | 82,728 | 61,765 |
| Média           | 10,250  | 10,668  | 8,134  | 5,919  |
| Desvio-Padrão   | 8,733   | 6,279   | 4,816  | 3,417  |
| RMS             | 13,466  | 12,379  | 9,453  | 6,835  |

População para 200 km<sup>2</sup>

| Estatística (m) | BD      | BS      | DH     | IV     |
|-----------------|---------|---------|--------|--------|
| Mínimo          | 0,002   | 0,233   | 0,190  | 0,118  |
| Máximo          | 169,000 | 140,076 | 82,728 | 61,765 |
| Média           | 10,156  | 10,637  | 8,105  | 5,899  |
| Desvio-Padrão   | 8,633   | 6,259   | 4,800  | 3,412  |
| RMS             | 13,329  | 12,341  | 9,420  | 6,815  |

População para 225 km<sup>2</sup>

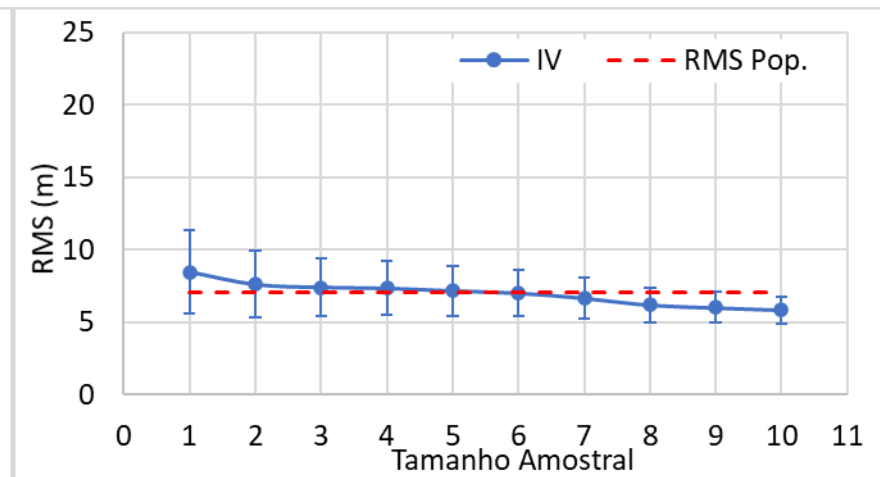
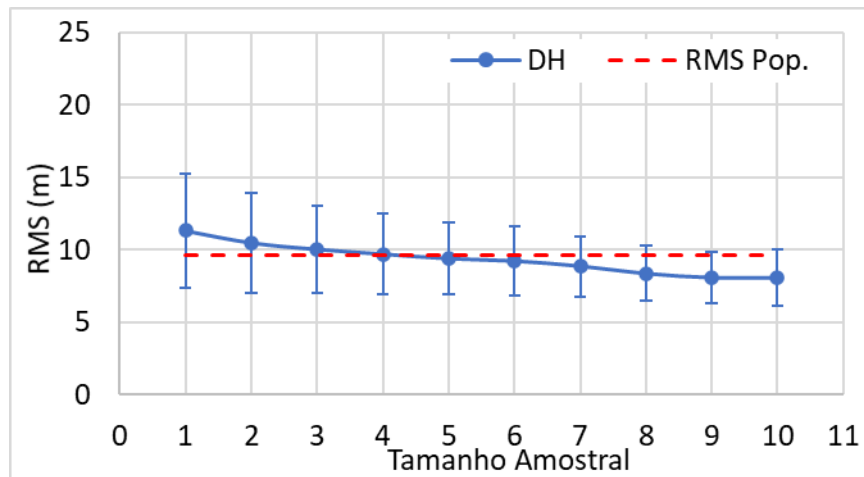
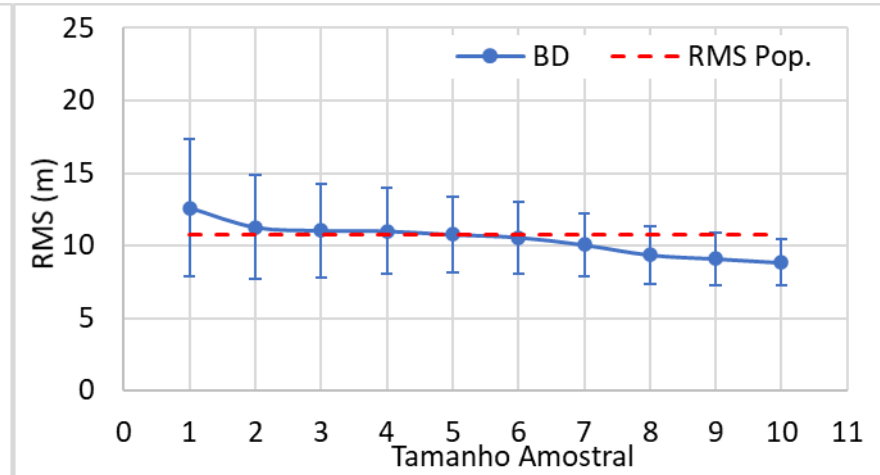
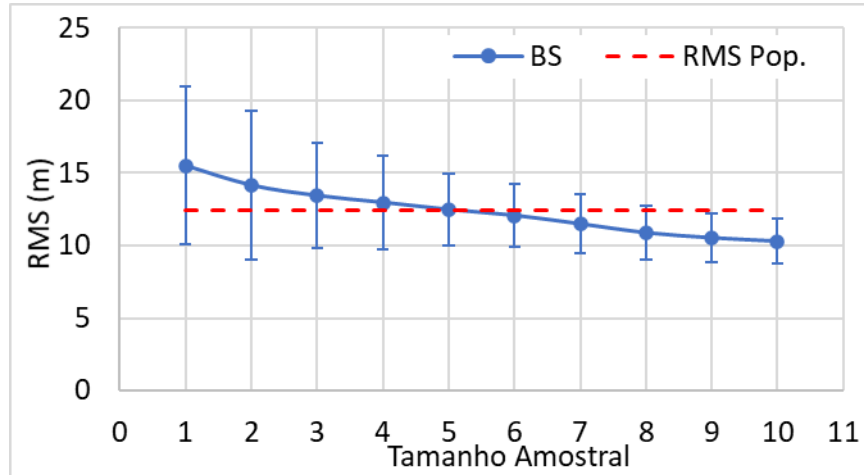
| Estatística (m) | BD      | BS      | DH      | IV     |
|-----------------|---------|---------|---------|--------|
| Mínimo          | 0,002   | 0,233   | 0,183   | 0,118  |
| Máximo          | 223,006 | 140,076 | 167,525 | 61,765 |
| Média           | 10,154  | 10,621  | 8,063   | 5,862  |
| Desvio-Padrão   | 8,722   | 6,388   | 4,981   | 3,388  |
| RMS             | 13,386  | 12,394  | 9,478   | 6,770  |

População para 250 km<sup>2</sup>

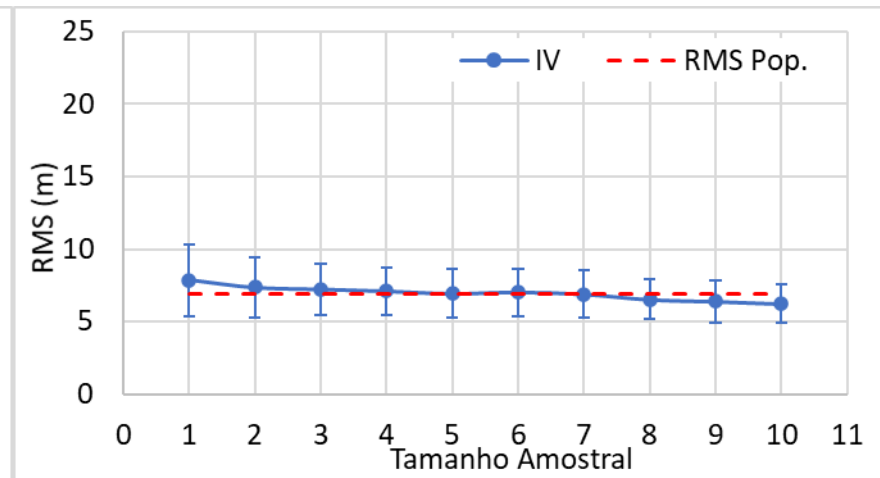
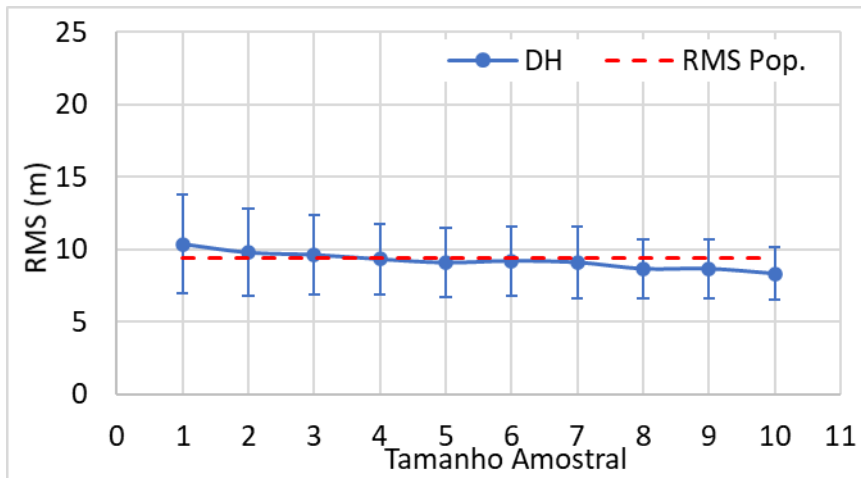
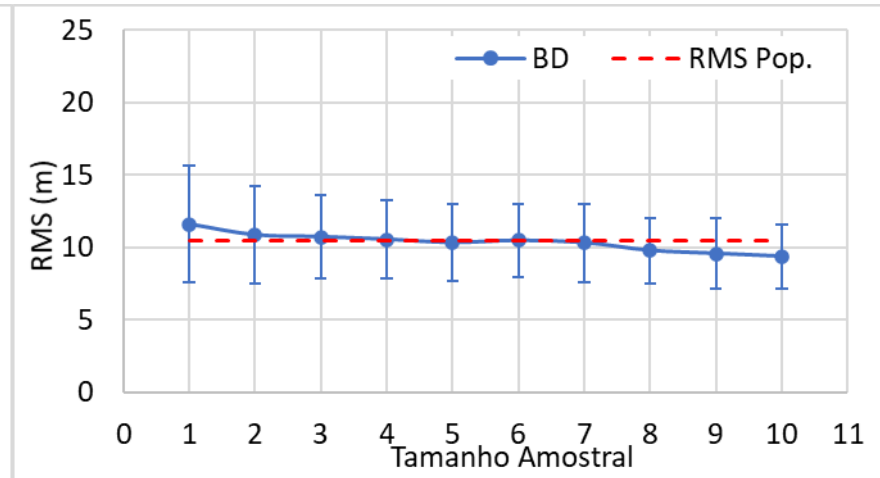
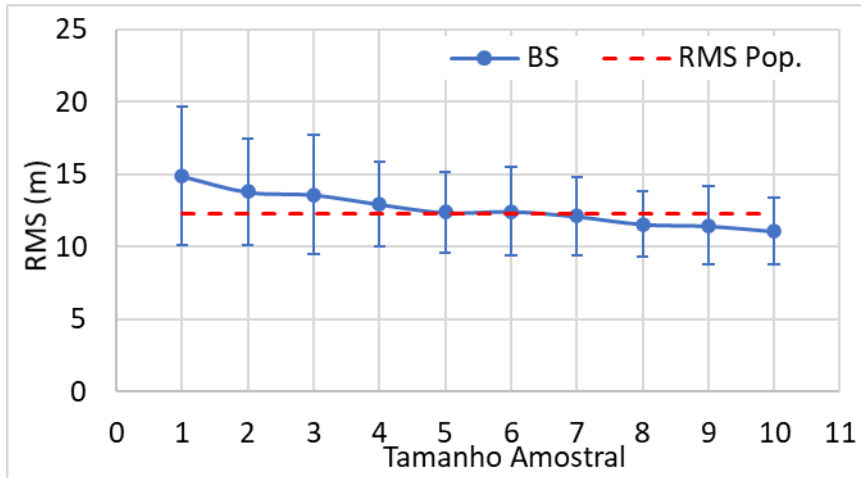
| Estatística (m) | BD      | BS      | DH      | IV     |
|-----------------|---------|---------|---------|--------|
| Mínimo          | 0,002   | 0,233   | 0,183   | 0,118  |
| Máximo          | 223,006 | 140,076 | 167,525 | 61,765 |
| Média           | 10,083  | 10,613  | 7,994   | 5,828  |
| Desvio-Padrão   | 8,599   | 6,338   | 4,928   | 3,345  |
| RMS             | 13,252  | 12,362  | 9,391   | 6,720  |

## Apêndice J. Relação entre o RMS e o Tamanho Amostral

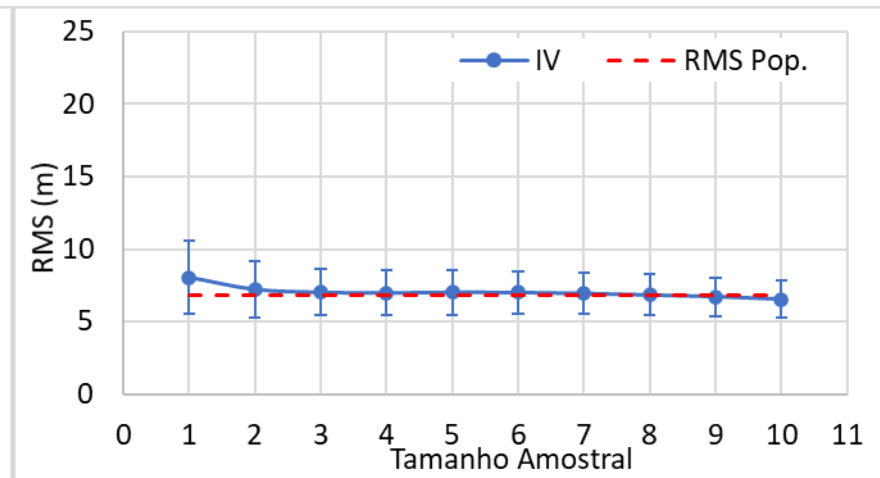
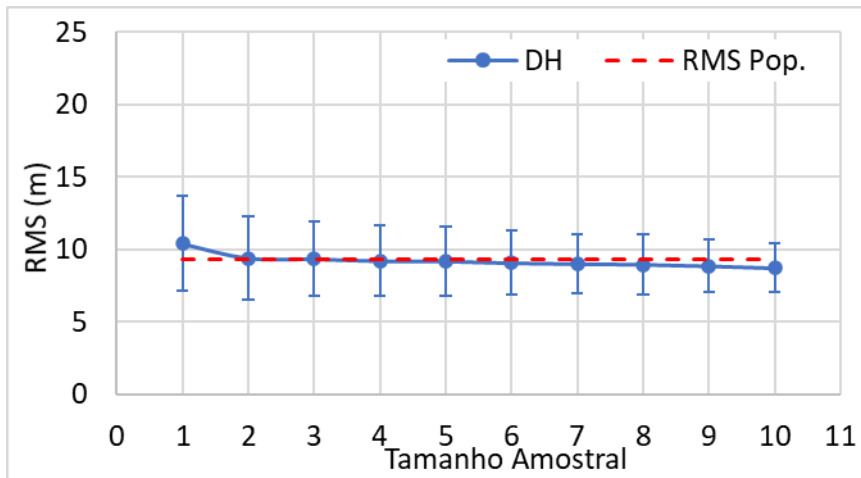
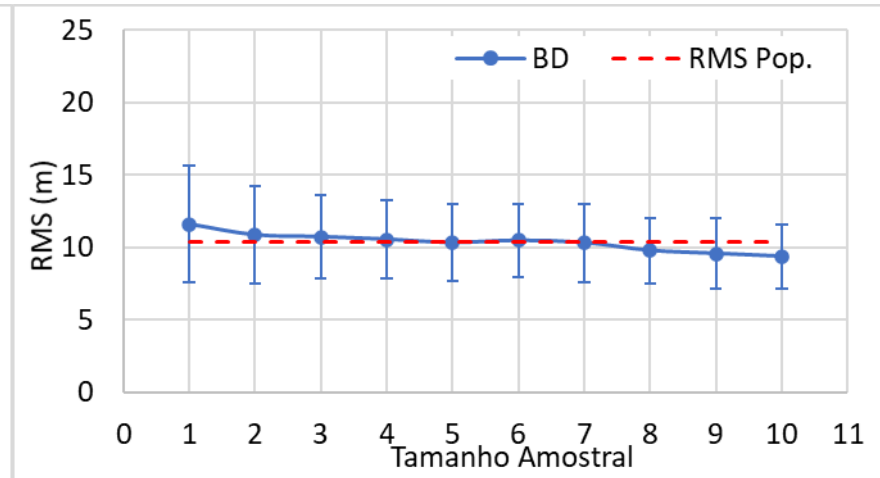
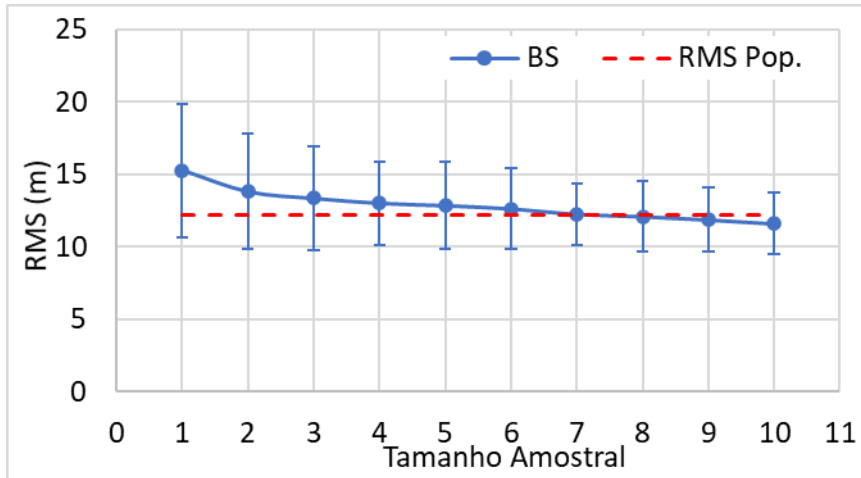
- Amostras em uma área de 5 km<sup>2</sup>



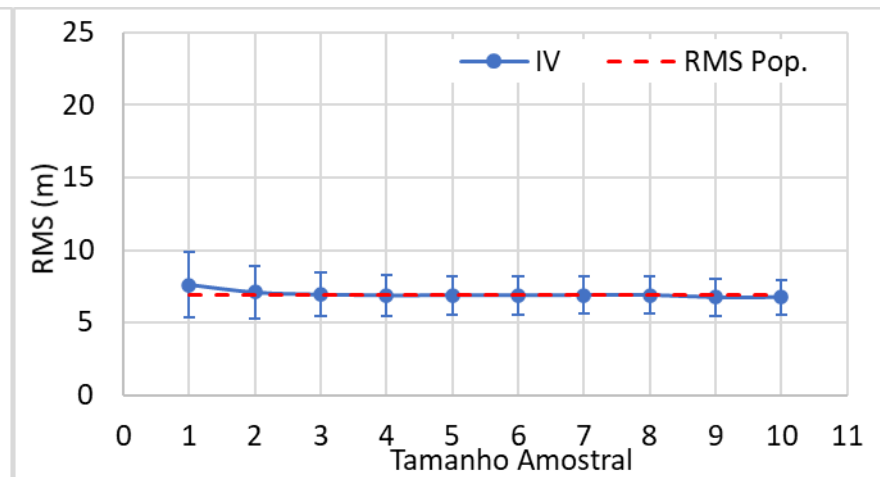
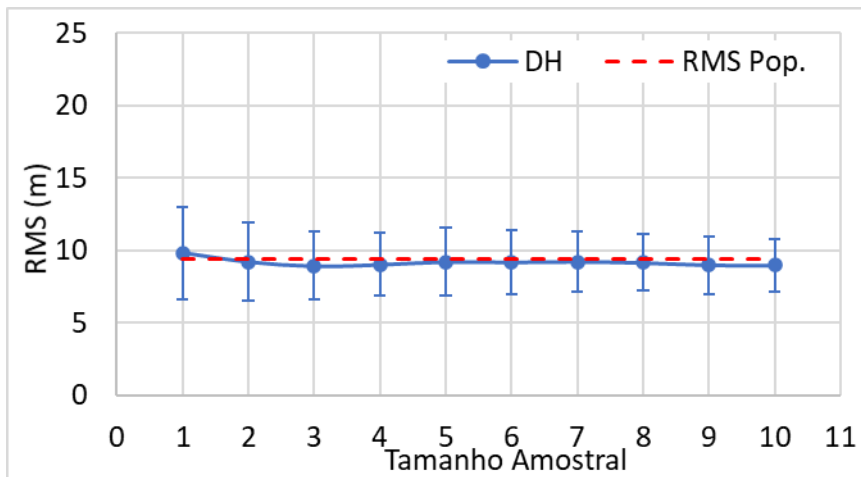
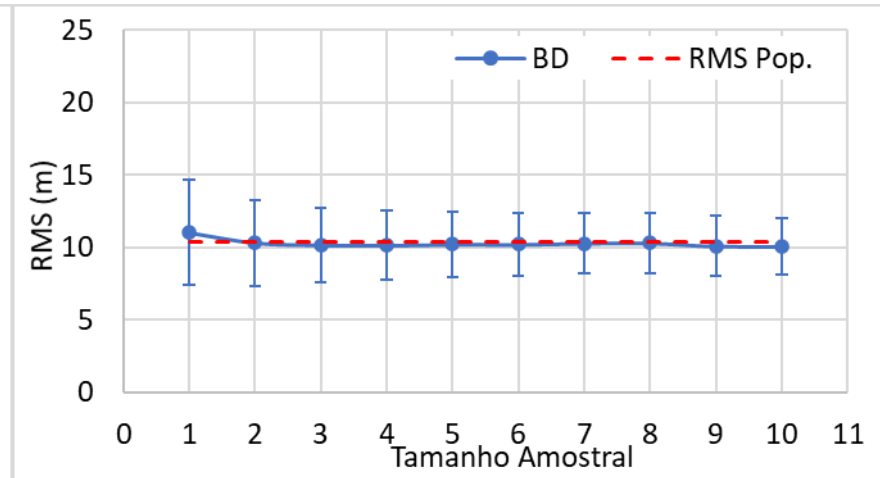
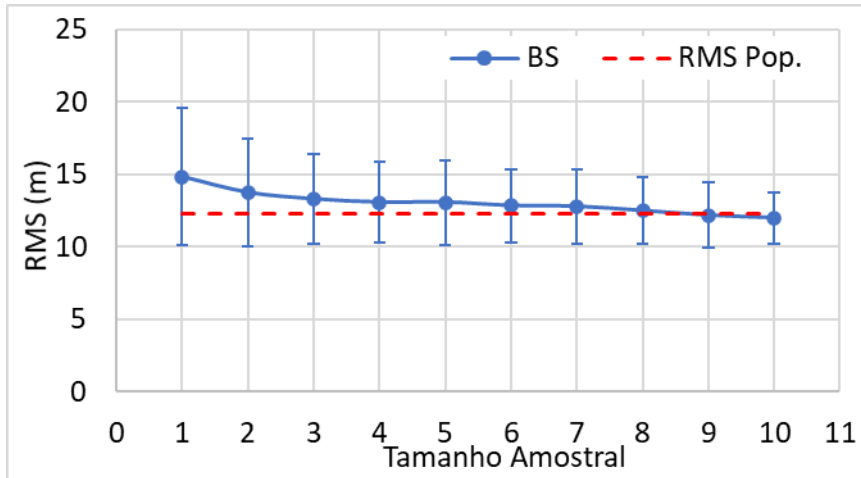
- Amostras em uma área de 7,5 km<sup>2</sup>



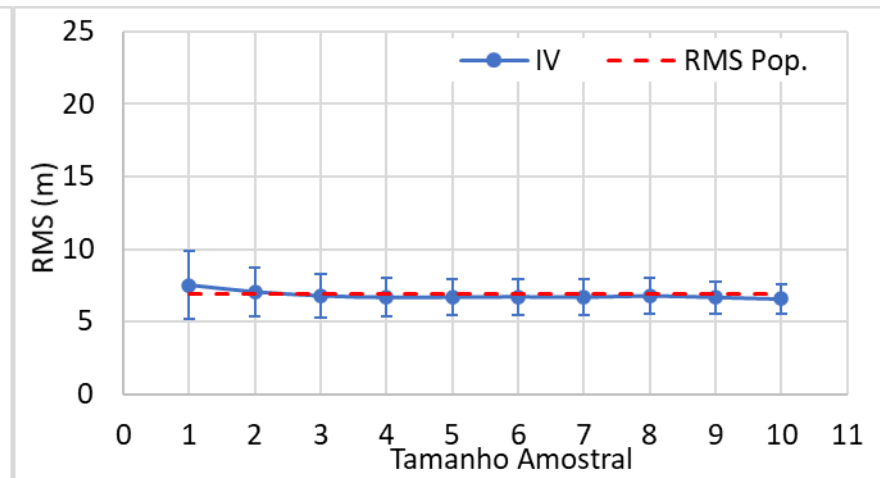
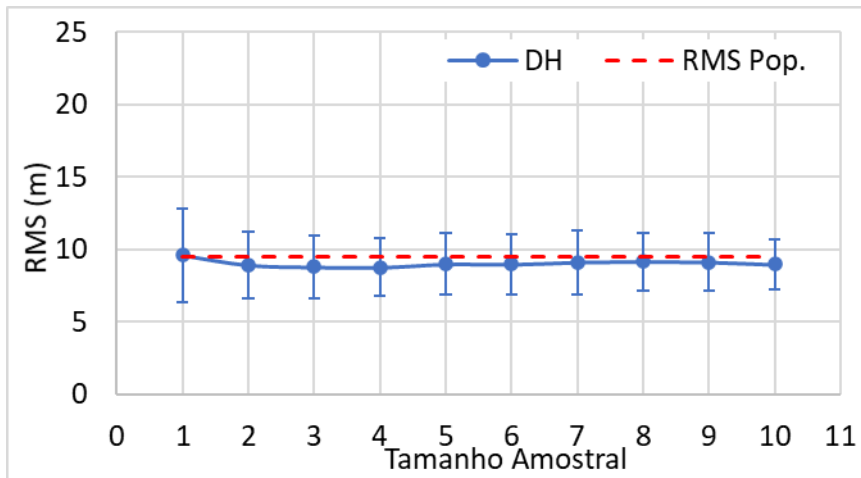
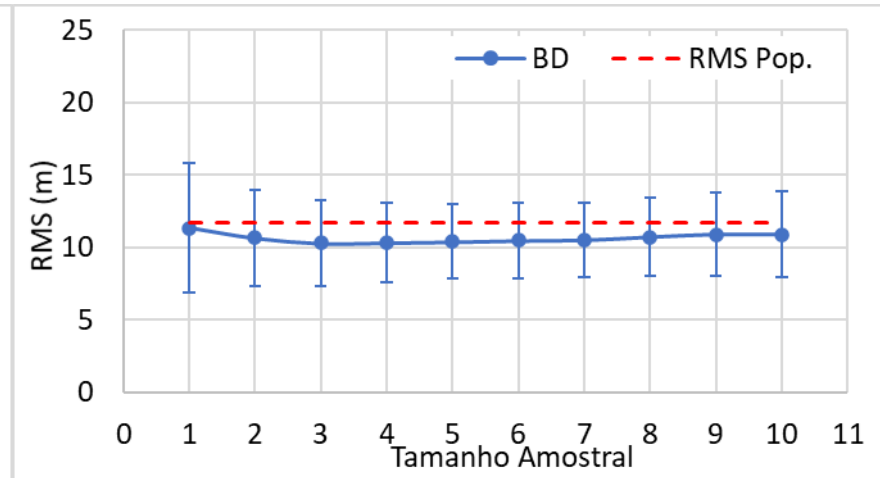
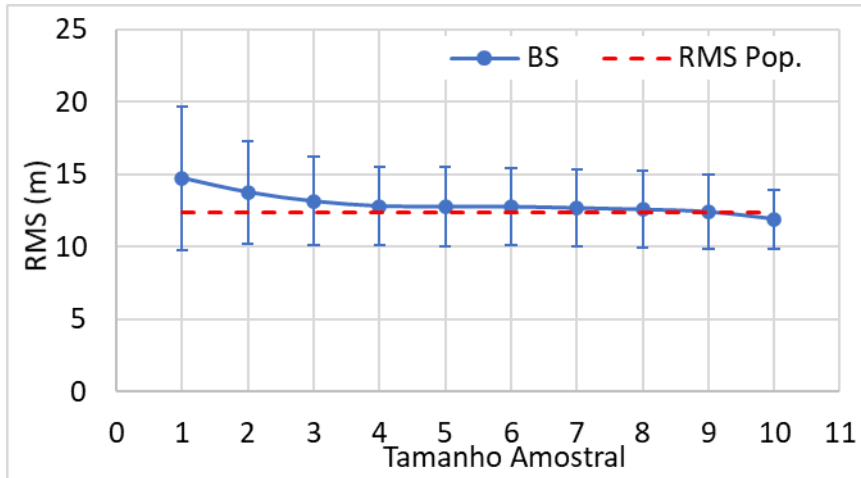
- Amostras em uma área de 10 km<sup>2</sup>



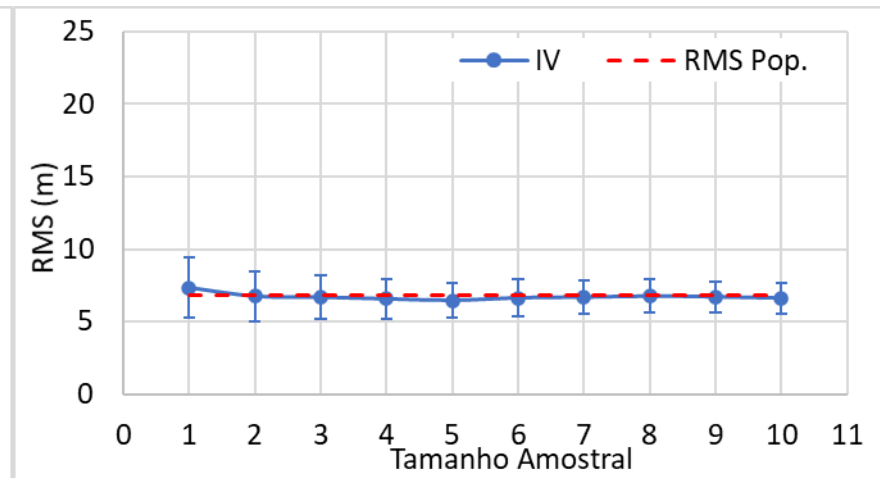
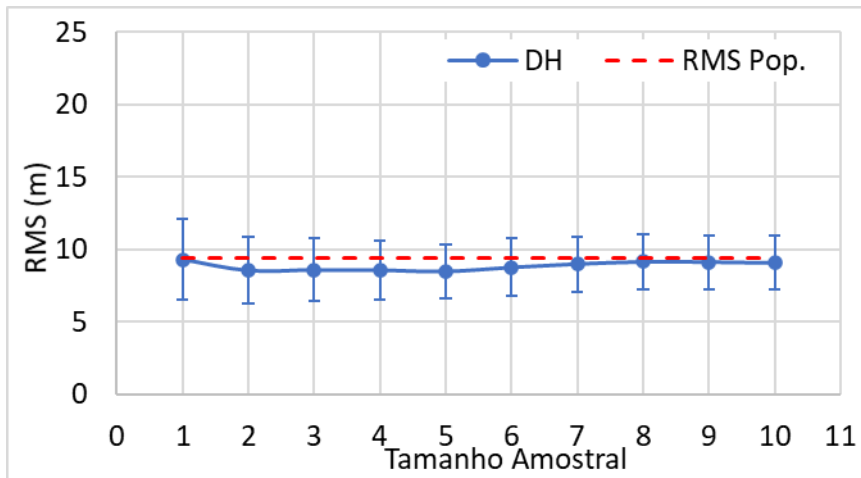
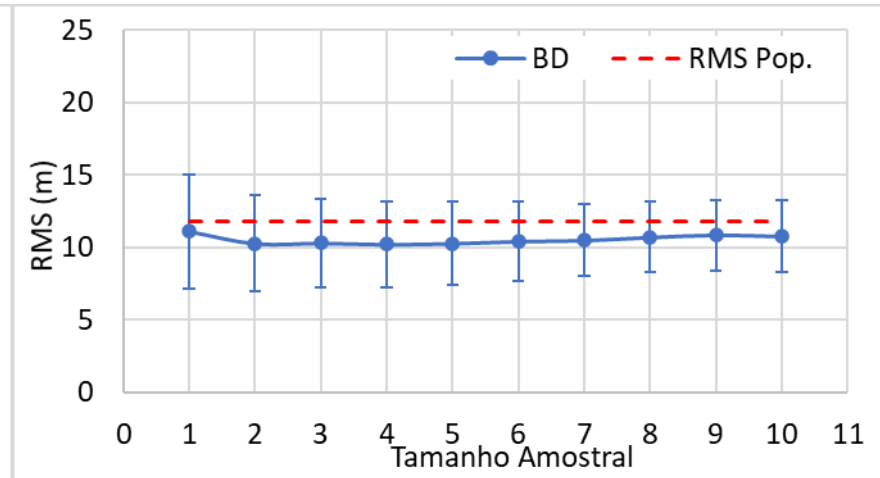
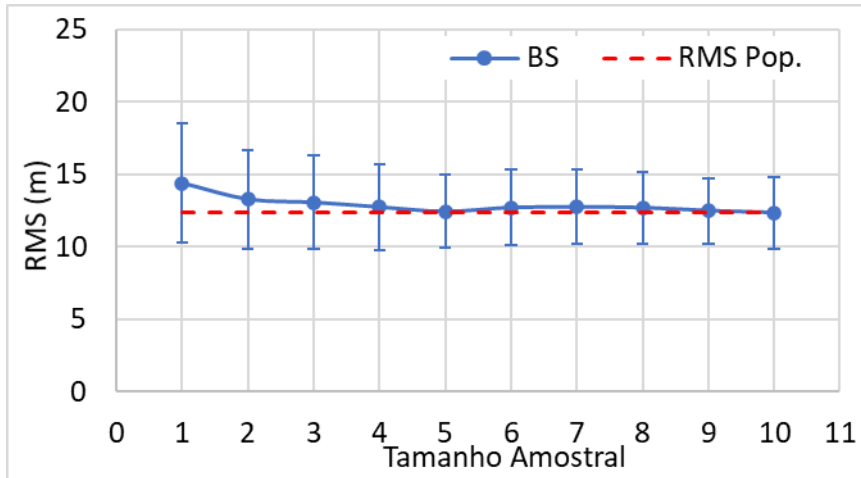
- Amostras em uma área de 15 km<sup>2</sup>



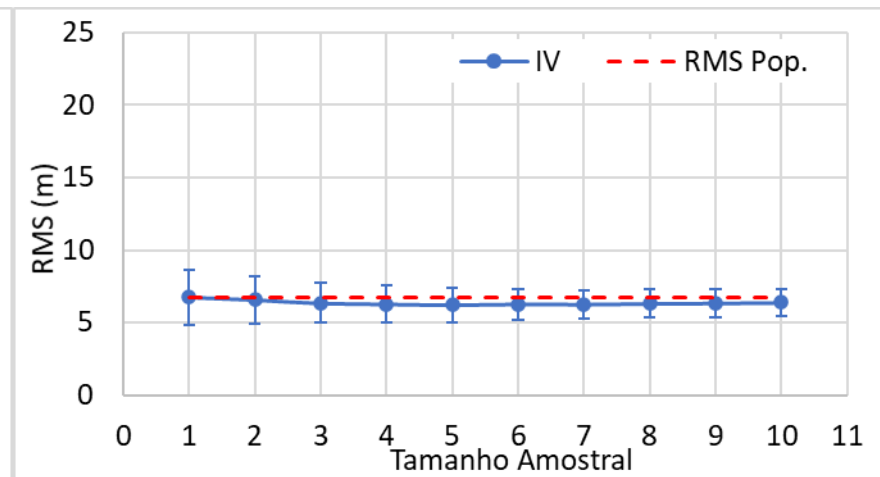
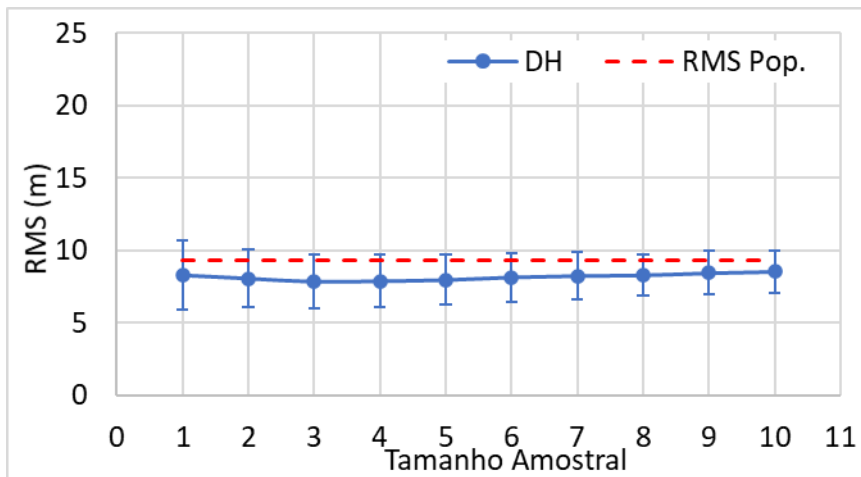
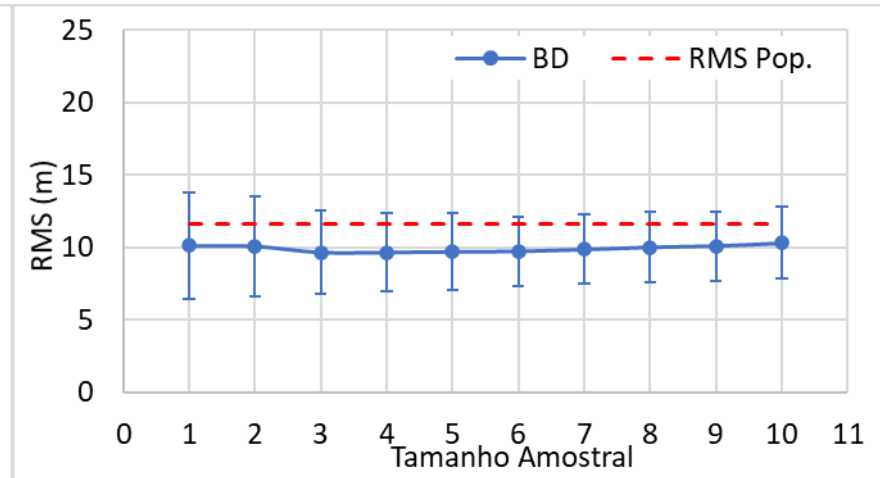
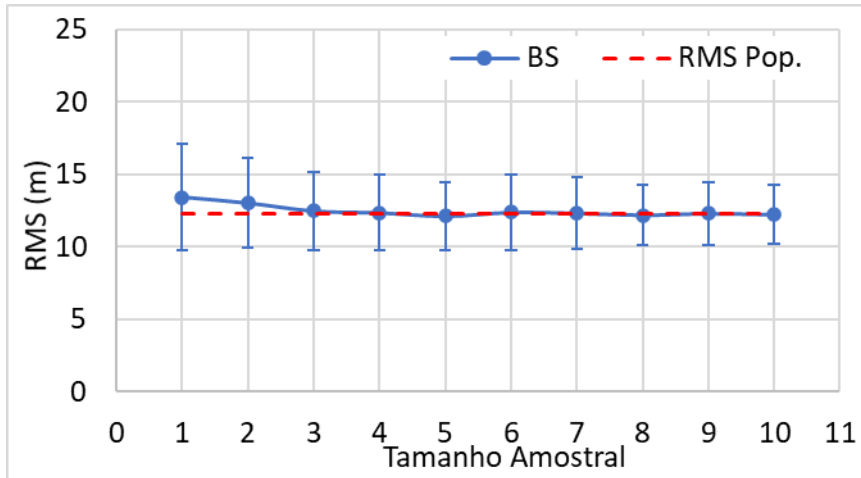
- Amostras em uma área de 20 km<sup>2</sup>



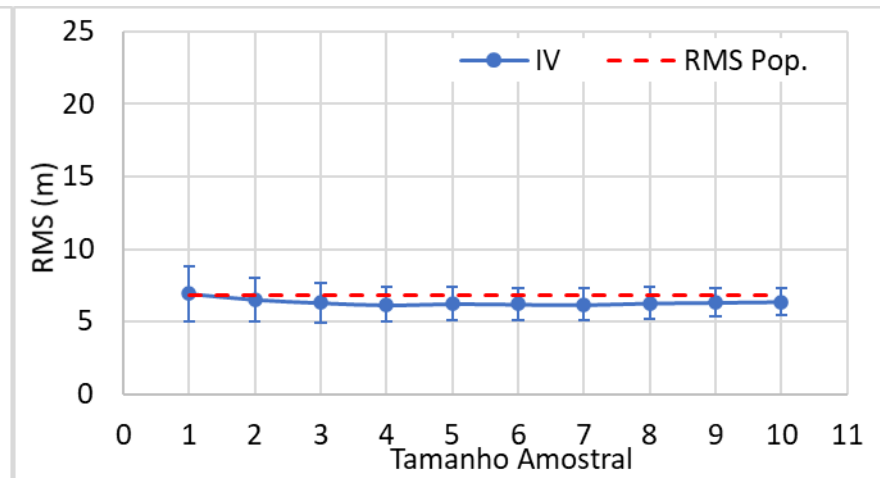
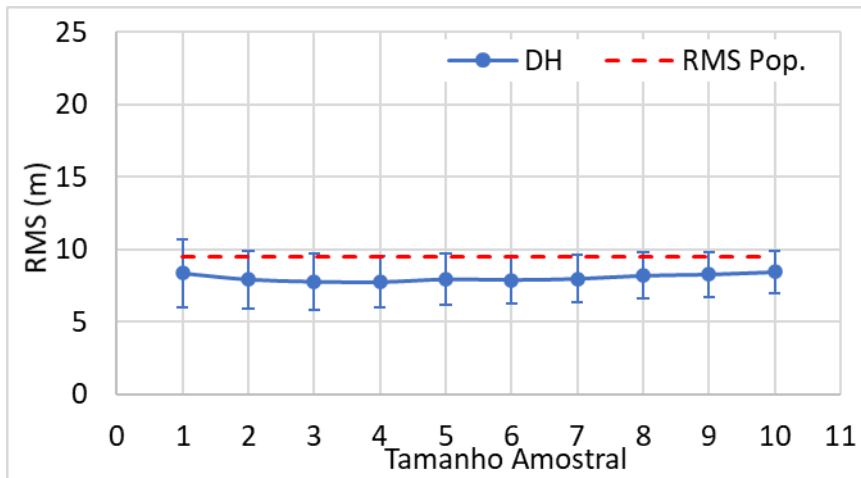
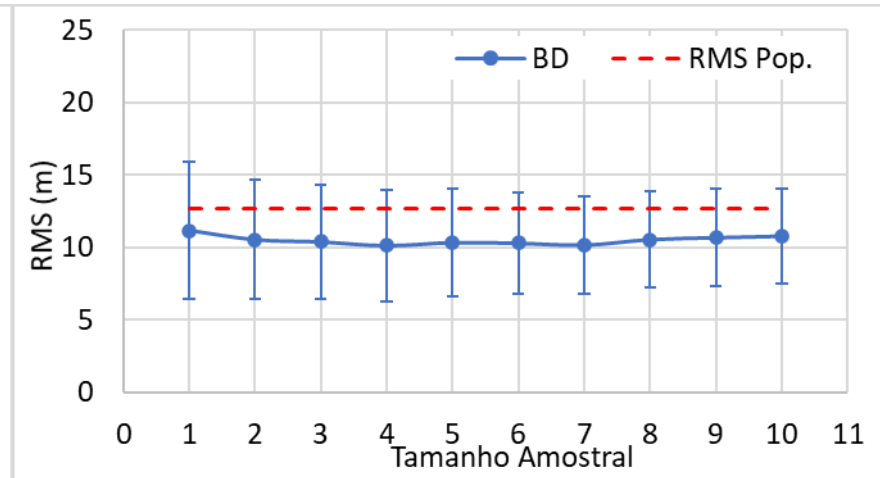
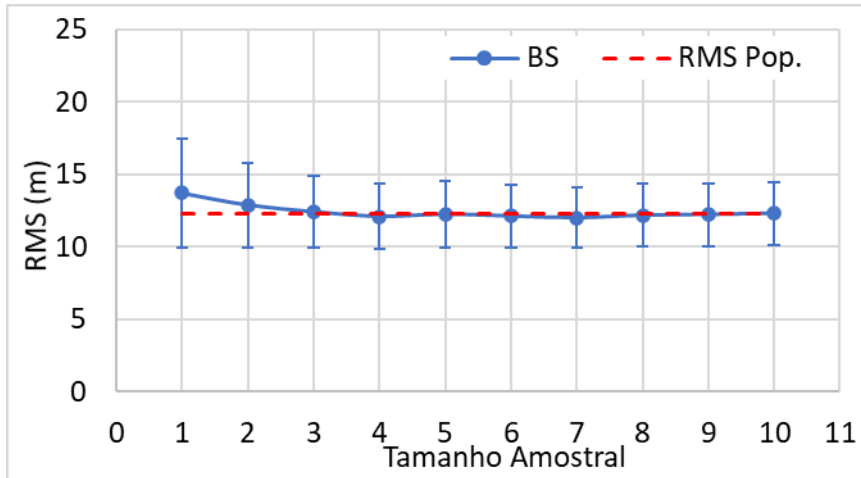
- Amostras em uma área de 25 km<sup>2</sup>



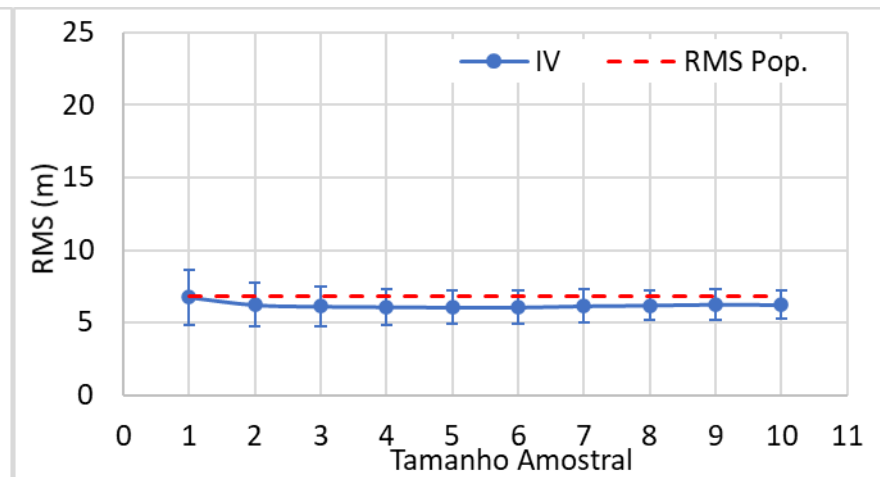
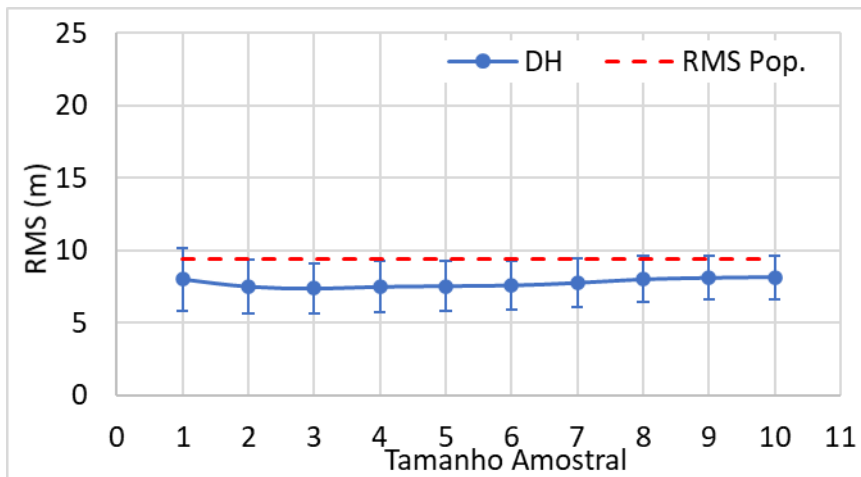
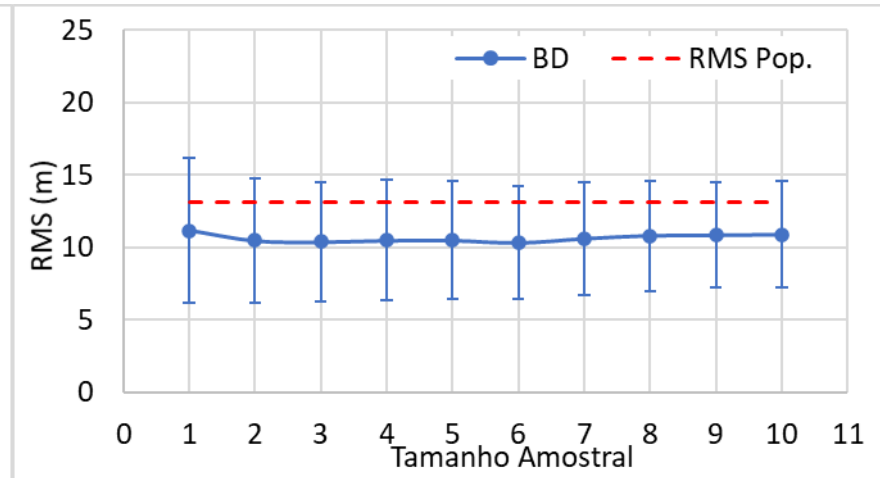
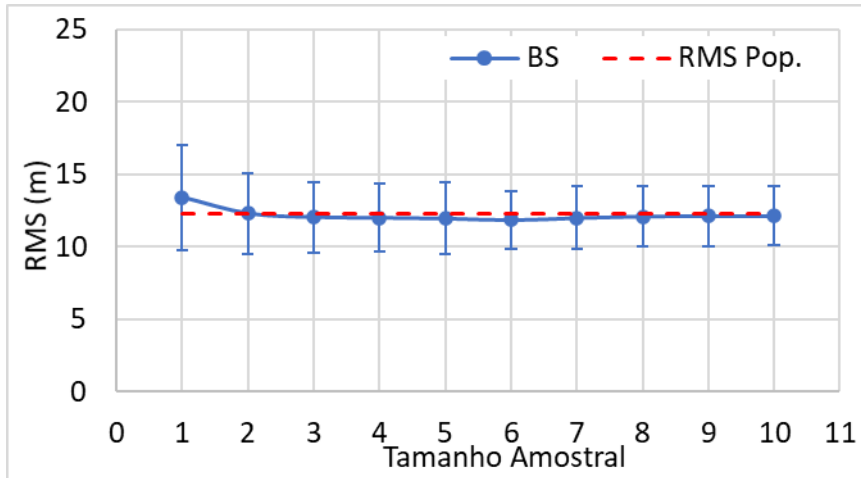
- Amostras em uma área de 50 km<sup>2</sup>



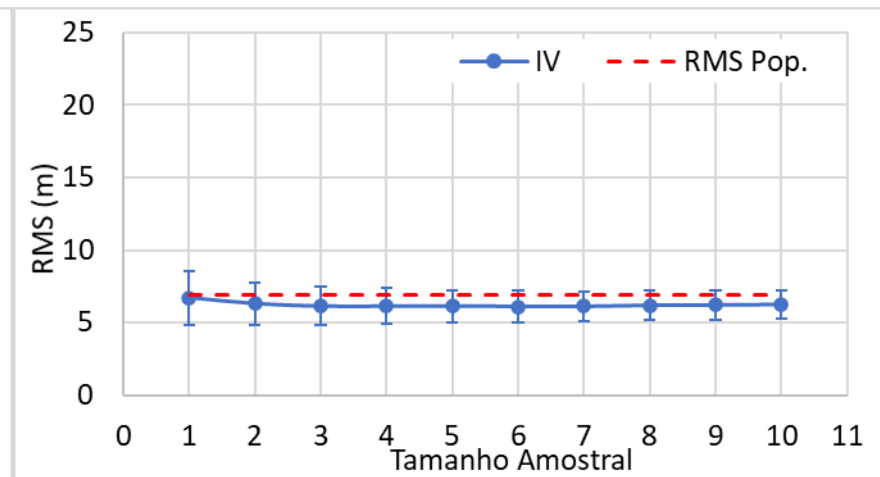
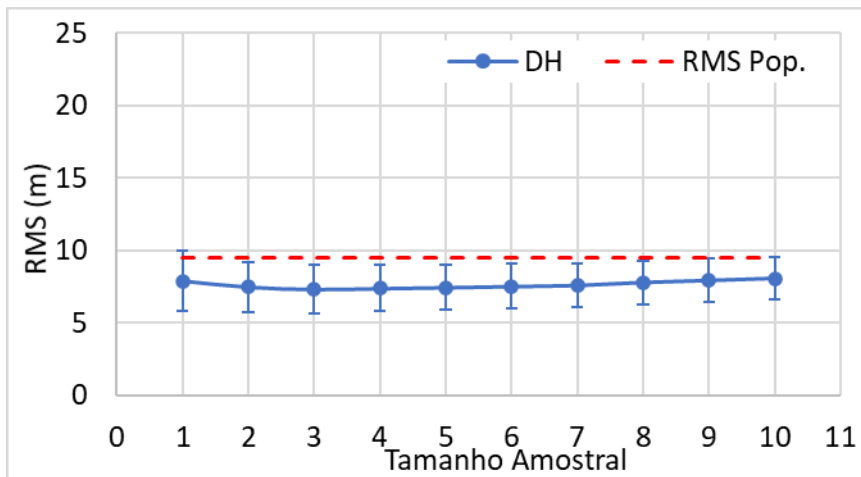
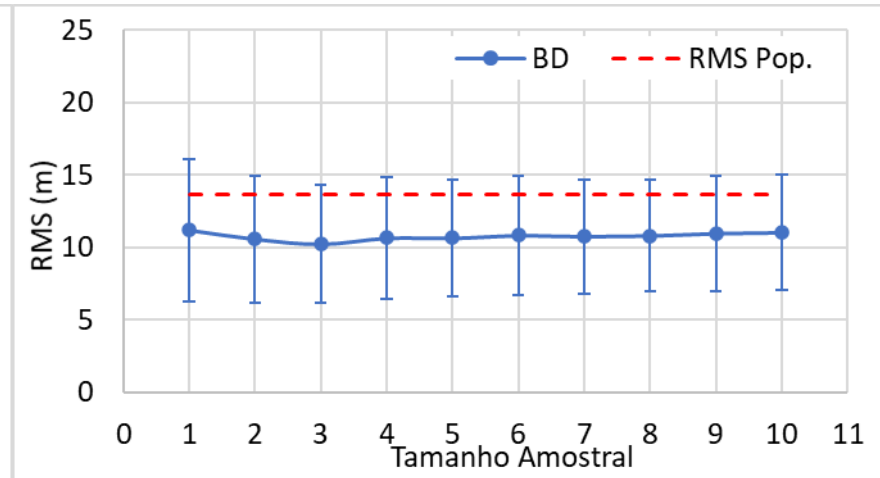
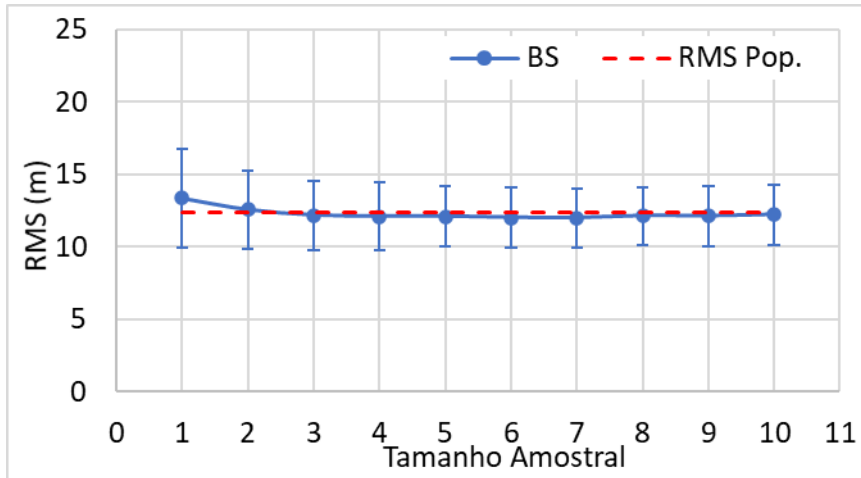
- Amostras em uma área de 75 km<sup>2</sup>



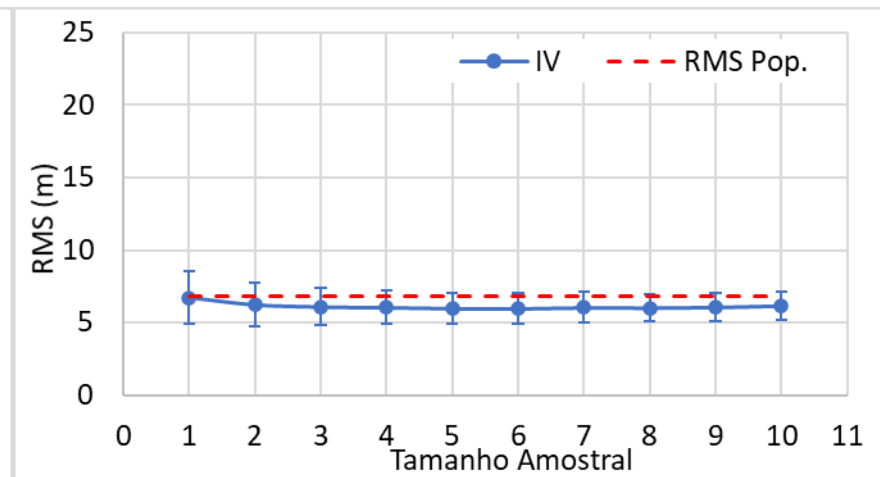
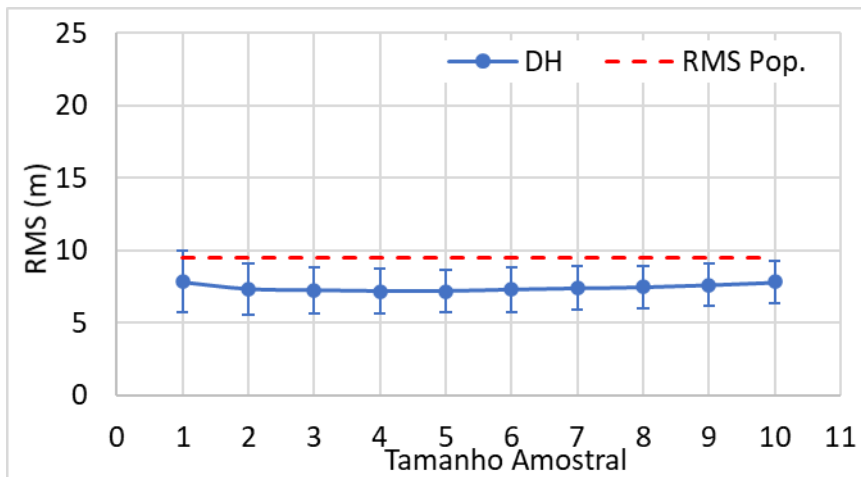
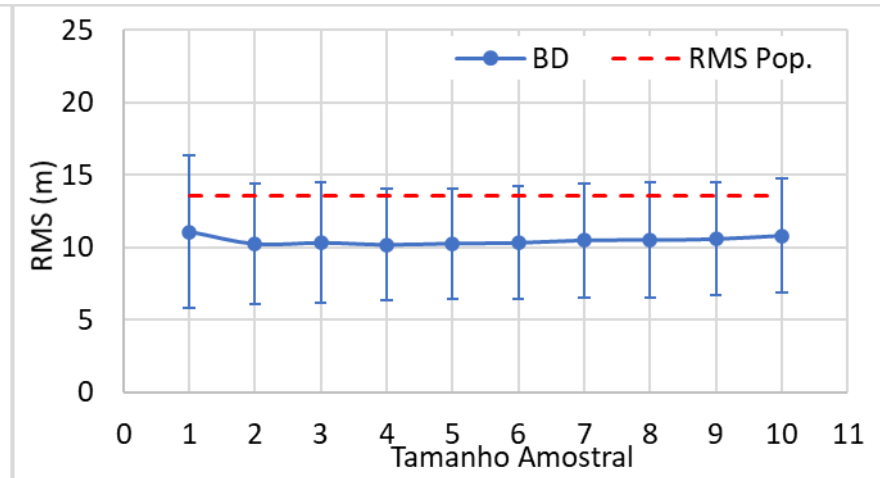
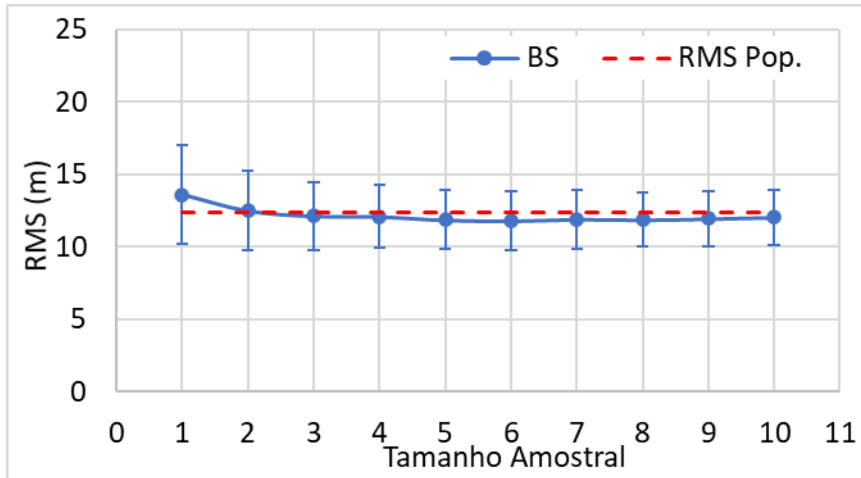
- Amostras em uma área de 100 km<sup>2</sup>



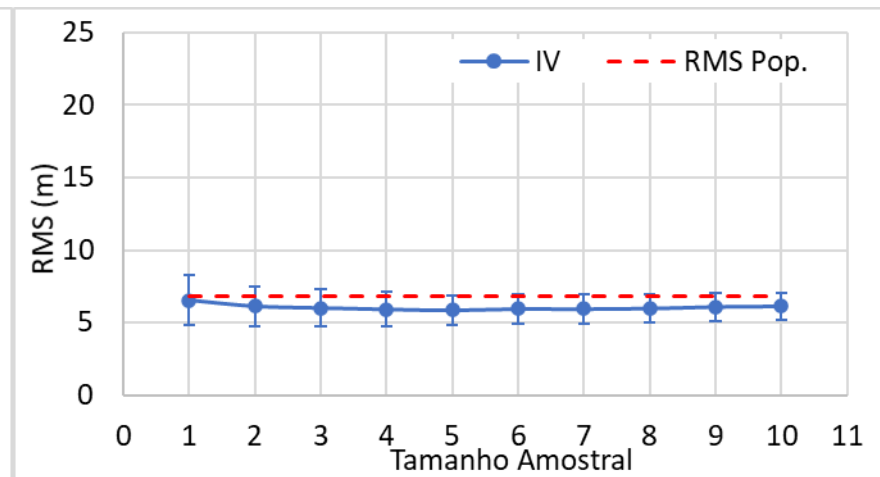
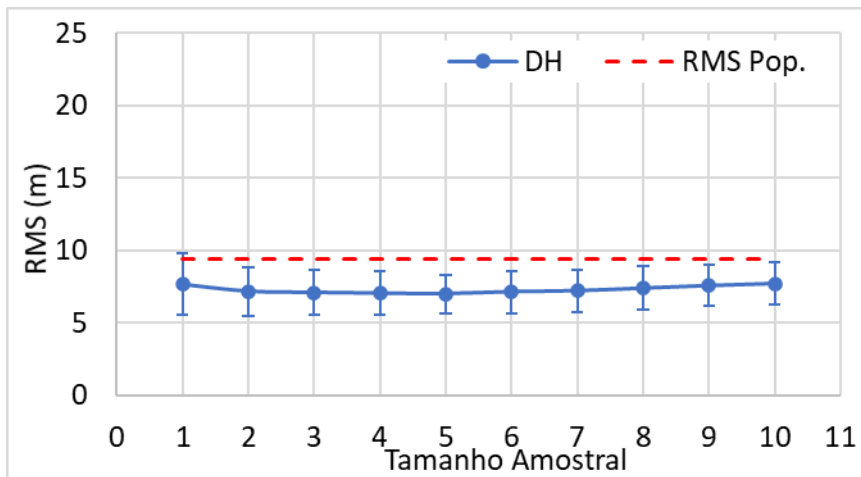
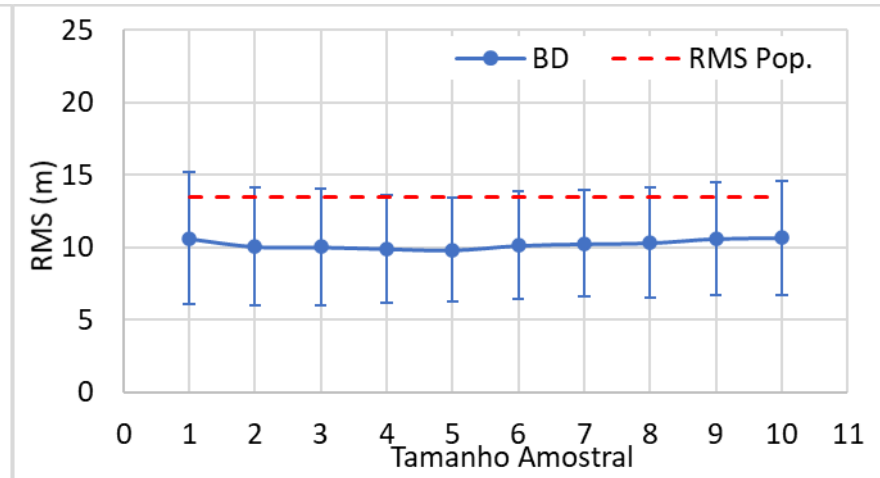
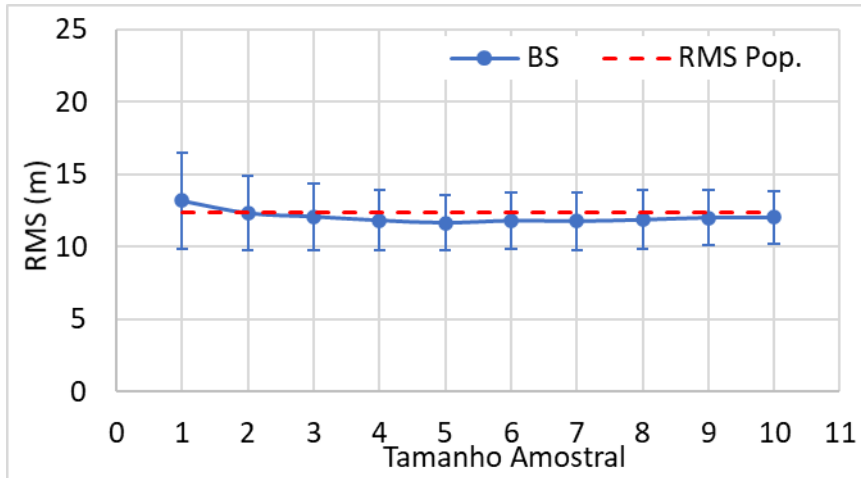
- Amostras em uma área de 125 km<sup>2</sup>



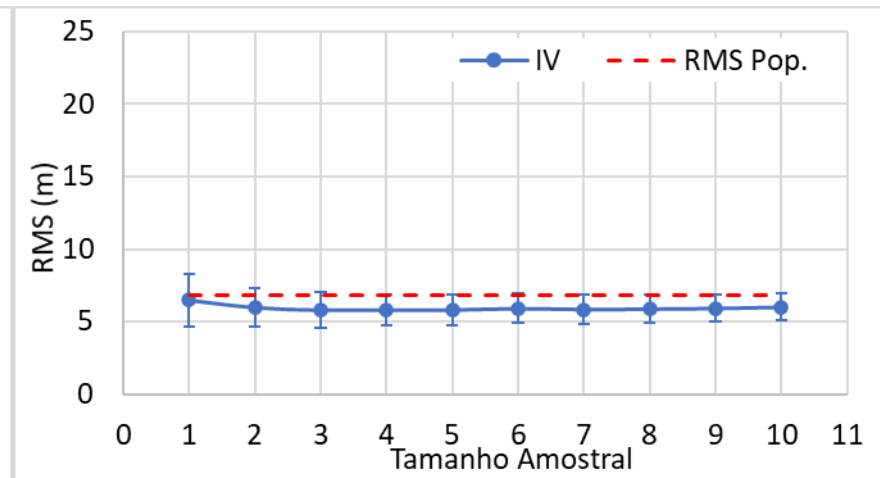
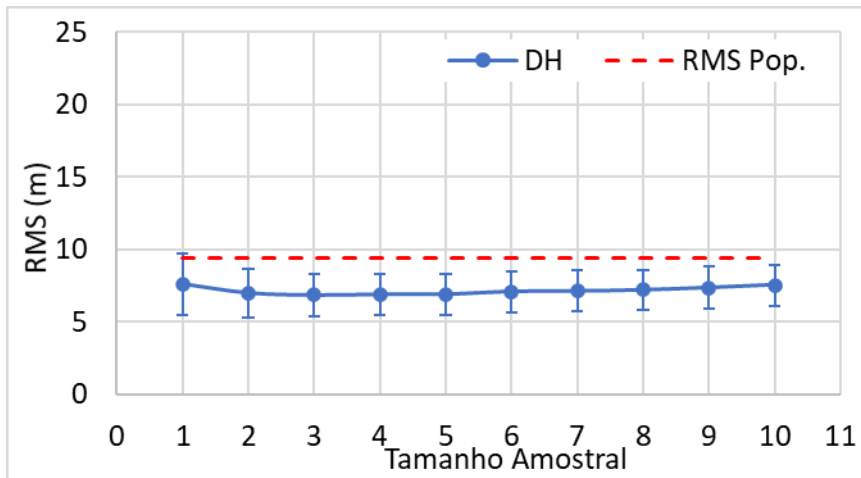
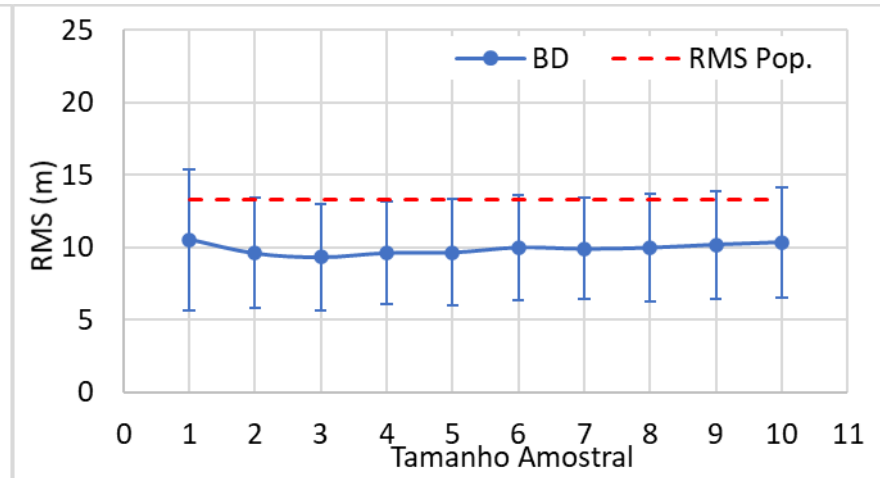
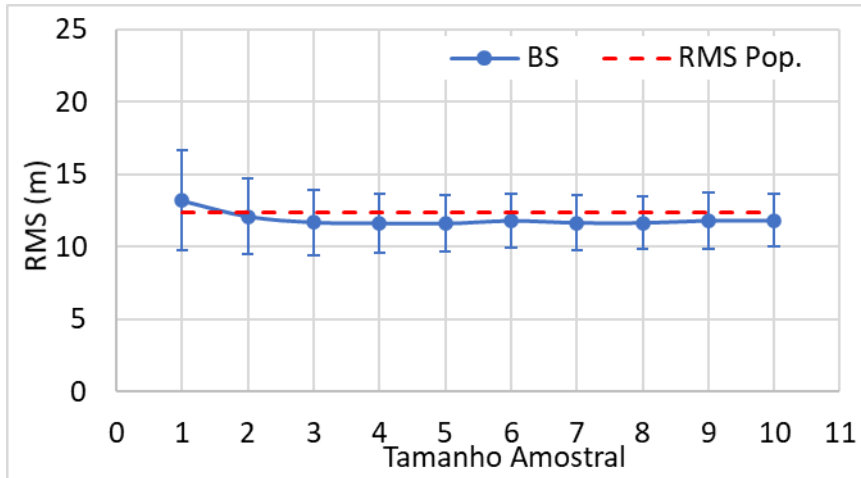
- Amostras em uma área de 150 km<sup>2</sup>



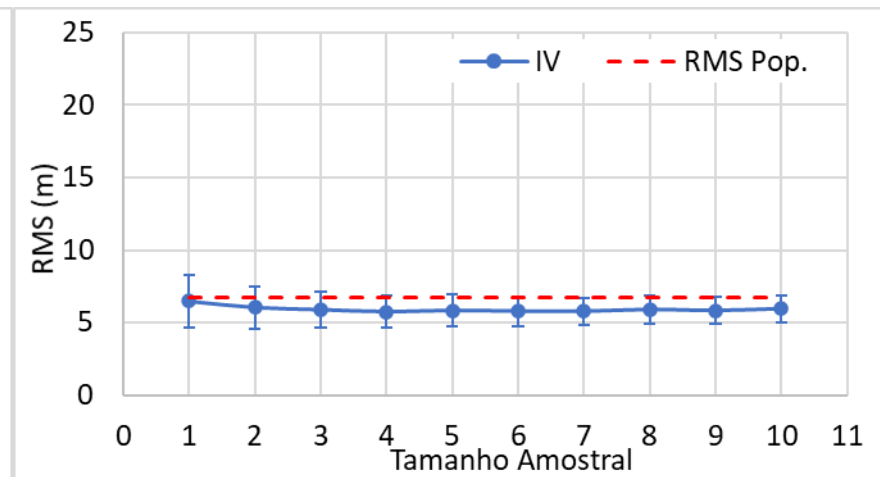
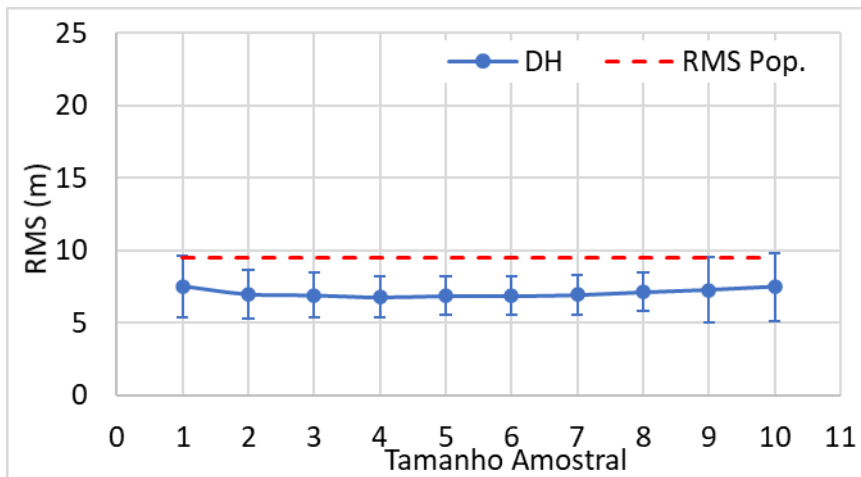
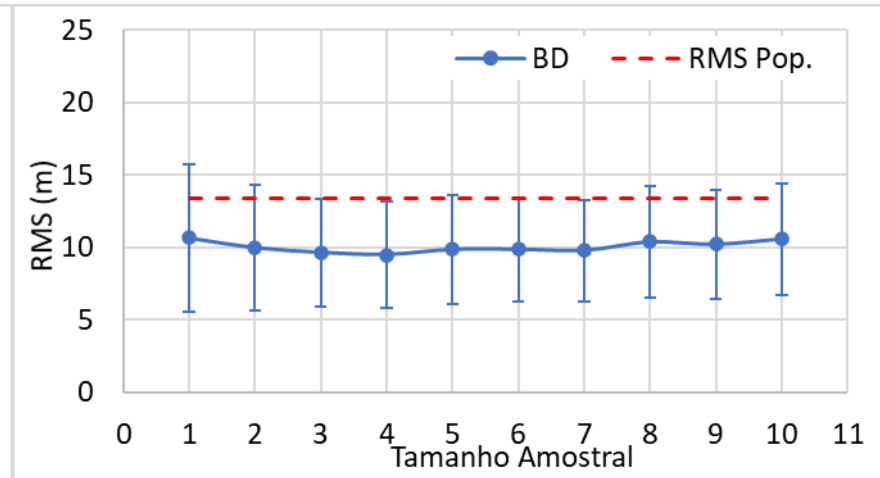
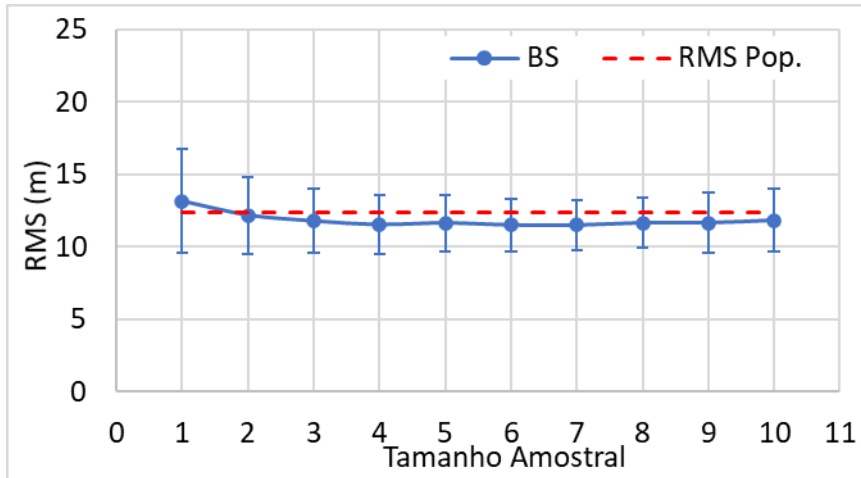
- Amostras em uma área de 175 km<sup>2</sup>



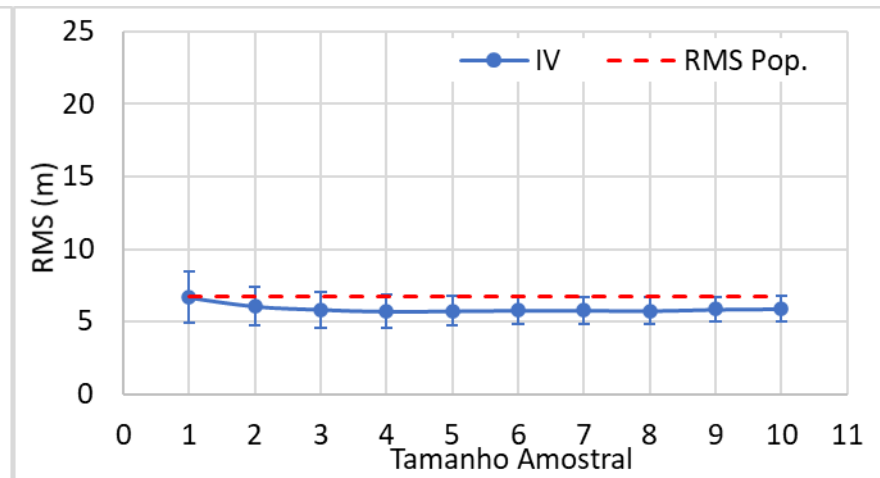
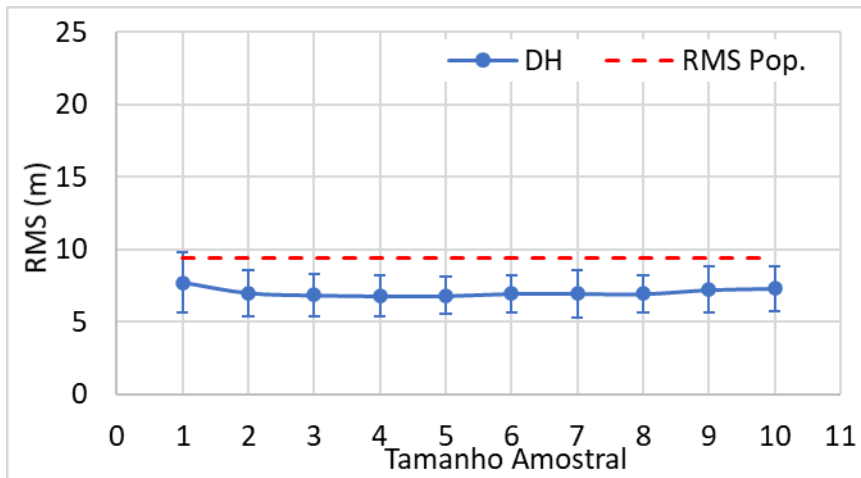
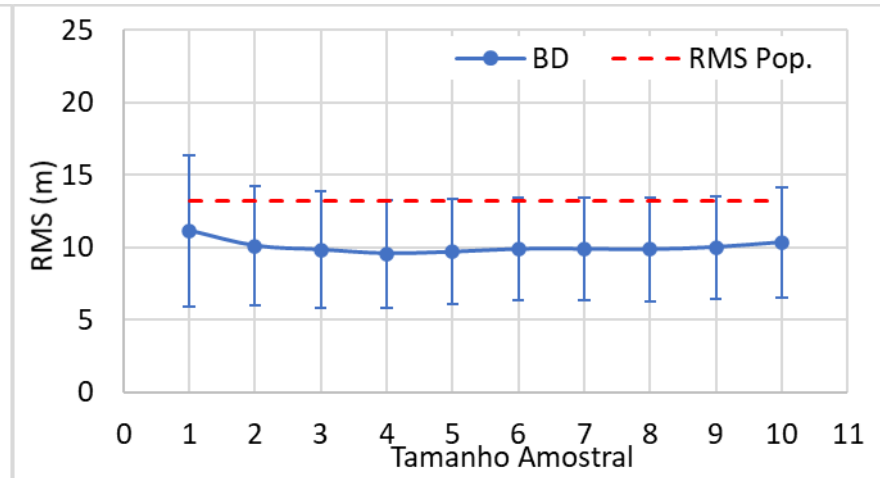
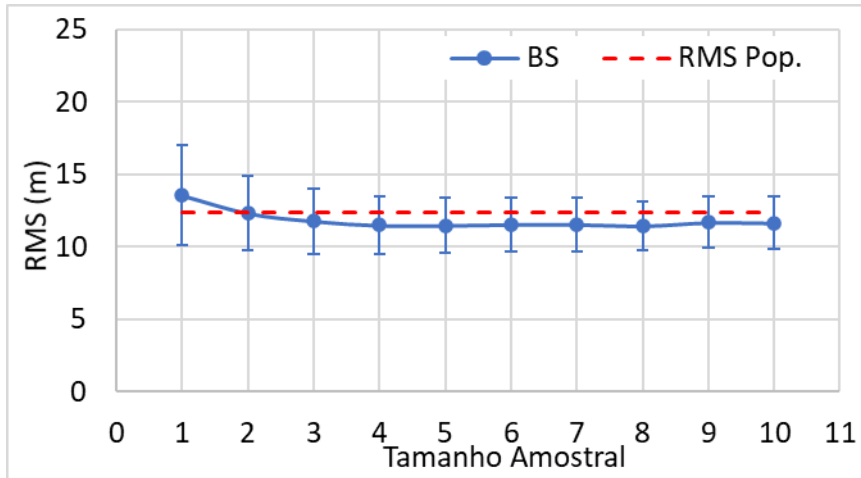
- Amostras em uma área de 200 km<sup>2</sup>



- Amostras em uma área de 225 km<sup>2</sup>



- Amostras em uma área de 250 km<sup>2</sup>



## Capítulo 3

### Apêndice A. Método TRS

# Método TRS (Translation, Rotation and Scale)

# Autora do método: Juliette Zanetti (ZANETTI, 2022)

# Autor do código: Marconi Martins Cunha

# OBSERVAÇÃO: Este código foi escrito para o caso em que as linhas de referência

# possuem mais vértices que as linhas de teste, que é o caso dessa

# pesquisa. Para o caso contrário, é necessário alterar a parte de

# segmentação das linhas

options(digits = 22)

# Carregar bibliotecas necessárias

library(raster)

library(sf)

library(magrittr)

library(dplyr)

library(maptools)

library(rgdal)

library(rgeos)

library(foreach)

library(doParallel)

library(sp)

library(ggplot2)

library(hrbrthemes)

```
library(lwgeom)
library(stats)
library(eixport)
library(SpatialGraph)
library(spatialEco)
library(spNetwork)
library(geos)
library(nngeo)
library(pracma)
library(plyr)

setwd("C:/...")

linhas_teste <- shapefile("...shp")
linhas_referencia <- shapefile("...shp")

#####

linhas_teste$ID_Unico <- linhas_teste$ID
linhas_referencia$ID_Unico <- linhas_referencia$ID

linhas_teste$ID_unico <- linhas_teste$ID
linhas_referencia$ID_unicoa <- linhas_referencia$ID

#####

linhas_referencia_sf <- st_as_sf(linhas_referencia)
linhas_referencia_sf <- st_zm(linhas_referencia_sf, drop = TRUE, what = "ZM")
```

```

linhas_referencia <- as_Spatial(linhas_referencia_sf)

# Segmentação das linhas

# 00. Ordenar a direção de construção das linhas - Marconi M.

# 0. Criar um campo 'ID_Unico' para as linhas de teste e de referência - Marconi M.

# 1. Identificar qual linha (teste ou referência) possui o menor número de vértices

# 2. A menor linha deve ser dividida pela mesma quantidade de vértices e com
TAMANHOS PROPORCIONAIS

# Extrair os vértices das linhas de referencia

vertices_linhas_referencia <- as(linhas_referencia, "SpatialPointsDataFrame")

rownames(vertices_linhas_referencia@data) <- seq(from = 1, to =
length(vertices_linhas_referencia))

vertices_linhas_teste <- as(linhas_teste, "SpatialPointsDataFrame")

rownames(vertices_linhas_teste@data) <- seq(from = 1, to =
length(vertices_linhas_teste))

### Preparação dos dados

# Extrair as coordenadas x e y dos pontos espaciais

X <- vector(mode = "double", length = dim(vertices_linhas_referencia)[1])
Y <- vector(mode = "double", length = dim(vertices_linhas_referencia)[1])

vertices_linhas_referencia$X <- X

```

```

vertices_linhas_referencia$Y <- Y

for(i in 1:dim(vertices_linhas_referencia)[1]){
  vertices_linhas_referencia@data[i,]$X <-
st_coordinates(sf::st_as_sf(vertices_linhas_referencia[i,]), "X")
  vertices_linhas_referencia@data[i,]$Y <-
st_coordinates(sf::st_as_sf(vertices_linhas_referencia[i,]), "Y")
}

# Criar um campo sequencial antes, para identificar cada vértice das linhas de
referência

vertices_linhas_referencia$seq <- ave(vertices_linhas_referencia$ID_Unico,
vertices_linhas_referencia$ID_Unico, FUN = seq_along)

# Campo da distância entre os vértices

Dist_vert <- vector(mode = "double", length = dim(vertices_linhas_referencia)[1])
vertices_linhas_referencia$Dist_vert <- Dist_vert

lista_subset <- list()
for (j in 1:dim(linhas_referencia)[1]) {
  vertices_linhas_referencia_subset <-
vertices_linhas_referencia[vertices_linhas_referencia$ID_Unico == j,]
  for(i in 1:(dim(vertices_linhas_referencia_subset)[1] - 1)){
    vertices_linhas_referencia_subset$Dist_vert[i + 1] <-
sqrt((vertices_linhas_referencia_subset[vertices_linhas_referencia_subset$seq
i,]@data$X -

```

```
vertices_linhas_referencia_subset[vertices_linhas_referencia_subset$seq == (i +
1),]@data$X)^2 +
```

```
(vertices_linhas_referencia_subset[vertices_linhas_referencia_subset$seq ==
i,]@data$Y
-
vertices_linhas_referencia_subset[vertices_linhas_referencia_subset$seq == (i +
1),]@data$Y)^2)
```

```
}
```

```
lista_subset[[j]] <- vertices_linhas_referencia_subset
```

```
}
```

```
if(length(lista_subset) > 1){
```

```
  vertices_linhas_referencia <- do.call(rbind, lista_subset)
```

```
} else {
```

```
  vertices_linhas_referencia <- lista_subset[[1]]
```

```
}
```

```
# Campo da distância ACUMULADA entre os vértices
```

```
Dist_vert_acc <- vector(mode = "double", length = dim(vertices_linhas_referencia)[1])
```

```
vertices_linhas_referencia$Dist_vert_acc <- Dist_vert_acc
```

```
lista_subset_2 <- list()
```

```
for (j in 1:dim(linhas_referencia)[1]) {
```

```
  vertices_linhas_referencia_subset_2 <-
```

```
vertices_linhas_referencia[vertices_linhas_referencia$ID_Unico == j,]
```

```
  for(i in 1:(dim(vertices_linhas_referencia_subset_2)[1] - 1)){
```

```

    vertices_linhas_referencia_subset_2$Dist_vert_acc[i + 1] <-
vertices_linhas_referencia_subset_2$Dist_vert_acc[i] +
vertices_linhas_referencia_subset_2$Dist_vert[i + 1]
}
lista_subset_2[[j]] <- vertices_linhas_referencia_subset_2
}

```

```

if(length(lista_subset_2) > 1){
  vertices_linhas_referencia <- do.call(rbind, lista_subset_2)
} else {
  vertices_linhas_referencia <- lista_subset_2[[1]]
}

```

# Campo da distância ACUMULADA PROPORCIONAL entre os vértices

```

Dist_acc_prop <- vector(mode = "double", length = dim(vertices_linhas_referencia)[1])
vertices_linhas_referencia$Dist_acc_prop <- Dist_acc_prop

```

```

lista_subset_3 <- list()
for (j in 1:dim(linhas_referencia)[1]) {
  vertices_linhas_referencia_subset_3 <-
vertices_linhas_referencia[vertices_linhas_referencia$ID_Unico == j,]
  for(i in 1:(dim(vertices_linhas_referencia_subset_3)[1])){
    vertices_linhas_referencia_subset_3$Dist_acc_prop[i] <-
vertices_linhas_referencia_subset_3$Dist_vert_acc[i]/vertices_linhas_referencia_sub
set_3$Dist_vert_acc[dim(vertices_linhas_referencia_subset_3)[1]]
  }
}

```

```

lista_subset_3[[j]] <- vertices_linhas_referencia_subset_3
}

if(length(lista_subset_3) > 1){
  vertices_linhas_referencia <- do.call(rbind, lista_subset_3)
} else {
  vertices_linhas_referencia <- lista_subset_3[[1]]
}

# Campo da distância NA LINHA DE TESTE onde deve estar os NOVOS vértices
Dist_vert_novos <- vector(mode = "double", length =
dim(vertices_linhas_referencia)[1])
vertices_linhas_referencia$Dist_vert_novos <- Dist_vert_novos

lista_subset_4 <- list()
for (i in 1:dim(linhas_referencia)[1]) {
  vertices_linhas_referencia_subset_4 <-
vertices_linhas_referencia[vertices_linhas_referencia$ID_Unico == i,]
  vertices_linhas_referencia_subset_4$Dist_vert_novos <-
vertices_linhas_referencia_subset_4$Dist_acc_prop*gLength(linhas_teste[linhas_
teste$ID_Unico == i,], byid = TRUE)
  lista_subset_4[[i]] <- vertices_linhas_referencia_subset_4
}

if(length(lista_subset_4) > 1){
  vertices_linhas_referencia <- do.call(rbind, lista_subset_4)
} else {

```

```

vertices_linhas_referencia <- lista_subset_4[[1]]
}

# Novos vértices

# ID_novos_vert <- vector()
lista_novos_vert <- list()
for (i in 1:dim(linhas_referencia)[1]) {
  vertices_linhas_referencia_subset_5 <-
vertices_linhas_referencia[vertices_linhas_referencia$ID_Unico == i,]
  vertices_linhas_referencia_subset_5 <-
gInterpolate(linhas_teste[linhas_teste$ID_Unico == i,],
vertices_linhas_referencia_subset_5$Dist_vert_novos, normalized = FALSE)
  lista_novos_vert[[i]] <- vertices_linhas_referencia_subset_5
}
novos_vert <- do.call(rbind, lista_novos_vert)

novos_vert <- as(novos_vert,"SpatialPointsDataFrame")
novos_vert@data <- vertices_linhas_referencia@data
novos_vert@proj4string <- linhas_teste@proj4string
linhas_teste_sf <- st_as_sf(linhas_teste)

# Atualizar as coordenadas
novos_vert@data$X <- st_coordinates(sf::st_as_sf(novos_vert))[, "X"]
novos_vert@data$Y <- st_coordinates(sf::st_as_sf(novos_vert))[, "Y"]

novos_vert_sf <- st_as_sf(novos_vert)

```

```
split_nodes <- data.frame(novos_vert@data)
```

```
len <- vector()
```

```
for(i in 1:length(vertices_linhas_referencia$Dist_vert_novos)){
```

```
  if(i == 1){
```

```
    len[i] <- vertices_linhas_referencia$Dist_vert_novos[i]
```

```
  } else {
```

```
    len[i] <- vertices_linhas_referencia$Dist_vert_novos[i] -  
vertices_linhas_referencia$Dist_vert_novos[i - 1]
```

```
  }
```

```
}
```

```
split_nodes <- dplyr::mutate(split_nodes, len = len,
```

```
                           cumlen = cumsum(len), newID = NA) #####
```

```
list_coords <- list()
```

```
for (i in 1:dim(linhas_teste)[1]) {
```

```
  coords <-
```

```
<-
```

```
dplyr::select(data.frame(sf::st_coordinates(linhas_teste_sf[linhas_teste_sf$ID_Unico  
== i,])), X, Y)
```

```
  ID_Unico <- rep(i, times = dim(coords)[1])
```

```
  coords <- data.frame(coords, ID_Unico)
```

```
  coords <- dplyr::mutate(coords, len = sqrt(((X - (lag(X)))^2) + (((Y - (lag(Y)))^2))))
```

```
  coords[["len"]] = ifelse(is.na(coords[["len"]]), 0, coords[["len"]])
```

```
  coords = dplyr::mutate(coords, cumlen = cumsum(coords[["len"]]), newID = " ")
```

```

list_coords[[i]] <- coords
}

coords <- do.call(rbind, list_coords)

# Subset e ordenar colunas

coords <- data.frame(coords$X, coords$Y, coords$len, coords$cumlen,
coords$newID, coords$ID_Unico)

colnames(coords) <- c("X", "Y", "len", "cumlen", "newID", "ID_Unico")

split_nodes <- data.frame(split_nodes$X, split_nodes$Y, split_nodes$len,
split_nodes$cumlen, split_nodes$newID, split_nodes$ID_Unico)

names(split_nodes) <- c("X", "Y", "len", "cumlen", "newID", "ID_Unico")

total_pts <- data.frame(rbind(split_nodes[2:nrow(split_nodes),], coords))

list_total_pts <- list()
for (i in 1:dim(linhas_teste)[1]) {
  subset_total_pts <- total_pts[total_pts$ID_Unico == i,]
  total_pts02 <- subset_total_pts[order(subset_total_pts$cumlen),]
  list_total_pts[[i]] <- total_pts02
}

total_pts02 <- do.call(rbind, list_total_pts)

rownames(total_pts02) <- c(1:nrow(total_pts02)) #####

list_split_pts <- list()

```

```

for (i in 1:dim(linhas_teste)[1]) {
  subset_split_pts <- total_pts02[total_pts02$ID_Unico == i,]

  split_points <- data.frame(start_nID = c(1,which(is.na(subset_split_pts$newID))),
stop_nID = c(which(is.na(subset_split_pts$newID)), nrow(subset_split_pts)))

  split_points$FID <- c(1:nrow(split_points)) #####
  ID_Unico <- rep(i, times = dim(split_points)[1])
  split_points <- data.frame(split_points, ID_Unico)

  list_split_pts[[i]] <- split_points
}

split_points <- do.call(rbind, list_split_pts)

new_line <- function(start_stop, coords) {
  return(sf::st_linestring(as.matrix(coords[start_stop[1]:start_stop[2], c("X", "Y")]))))
}

list_split_lines <- list()
for (i in 1:dim(linhas_teste)[1]) {
  total_pts02_sub <- total_pts02[total_pts02$ID_Unico == i,]
  split_points_sub <- split_points[split_points$ID_Unico == i,]

  subset_split_lines <- plyr::apply(as.matrix(split_points_sub[,c("start_nID",
"stop_nID")]),
                                .margins = 1, .fun = new_line, coords = total_pts02_sub)

  sf_lines <- st_sf(geometry = st_sfc(subset_split_lines, crs = sf::st_crs(str)))
  sf_lines$ID_Unico <- rep(i, times = dim(sf_lines)[1])

  list_split_lines[[i]] <- sf_lines
}

```

```

sf_lines <- do.call(rbind, list_split_lines)
sf_lines$length <- as.numeric(st_length(sf_lines))

# Excluir sf_lines com comprimento menor que 1 mm
sf_lines <- sf_lines[sf_lines$length > 0.001,]

vertices_linhas_referencia_sf <- st_as_sf(vertices_linhas_referencia)

# Criar um campo sequencial 'Seq_len' baseado na 'Dist_acc_prop'. Esse campo deve
ser usado

# para calcular o azimute
novos_vert_sf <- novos_vert_sf %>%
  group_by(ID_Unico) %>%
  mutate(Seq_len = dense_rank(Dist_acc_prop)) %>%
  ungroup()

vertices_linhas_referencia_sf <- vertices_linhas_referencia_sf %>%
  group_by(ID_Unico) %>%
  mutate(Seq_len = dense_rank(Dist_acc_prop)) %>%
  ungroup()

### Translação

Resultado_Translacao <- vector()

```

```

novos_vert_sf_list <- list()
for (j in 1:dim(linhas_teste)[1]) {
  novos_vert_sf_sub <- novos_vert_sf[novos_vert_sf$ID_Unico == j,]
  vertices_linhas_referencia_sf_sub <-
vertices_linhas_referencia_sf[vertices_linhas_referencia_sf$ID_Unico == j,]

  Az <- vector()
  ID <- vector()
  ID_Unico <- vector()
  for (i in 1:dim(novos_vert_sf_sub)[1]) {
    Az[i] <-
ngeo::st_azimuth(vertices_linhas_referencia_sf_sub[vertices_linhas_referencia_sf_
sub$seq == i,],
                  novos_vert_sf_sub[novos_vert_sf_sub$seq == i,])
    ID[i] <- novos_vert_sf_sub$seq[i]
    ID_Unico[i] <- j
  }
  Az.df <- data.frame(ID, Az, ID_Unico)
  names(Az.df) <- c("seq", "Az", "ID_Unico")
  novos_vert_sf_cbind <- cbind(novos_vert_sf_sub, Az.df)

  novos_vert_sf_list[[j]] <- novos_vert_sf_cbind # Az.df
}
novos_vert_sf_df <- do.call(rbind, novos_vert_sf_list)
novos_vert_sf <- novos_vert_sf_df

```

```

# Cálculo da Média Direcional

novos_vert_sf$Sen_az <- sin(pracma::deg2rad(novos_vert_sf$Az))
novos_vert_sf$Cos_az <- cos(pracma::deg2rad(novos_vert_sf$Az))

novos_vert_sf$Sum_Sen <- sum(novos_vert_sf$Sen_az)
novos_vert_sf$Sum_Cos <- sum(novos_vert_sf$Cos_az)

novos_vert_sf$Theta_Rad <-
atan(novos_vert_sf$Sum_Sen/novos_vert_sf$Sum_Cos)
novos_vert_sf$Theta_Deg <- pracma::rad2deg(novos_vert_sf$Theta_Rad)

if(novos_vert_sf$Theta_Deg[1] < 0){
  novos_vert_sf$Theta_Deg <- novos_vert_sf$Theta_Deg + 360
}

# Cálculo da Variância Circular

novos_vert_sf$Cr <- sqrt((sum(novos_vert_sf$Sen_az)^2 +
(sum(novos_vert_sf$Cos_az)^2)
novos_vert_sf$Sc <- 1 - novos_vert_sf$Cr[1]/dim(novos_vert_sf)[1]

# Se Sc > 0.5 -> Não há tendência
# Se Sc <= 0.5 -> Há tendência
if(novos_vert_sf$Sc[1] > 0.5){
  Resultado_Translacao <- c("Produto sem tendência de TRANSLAÇÃO")
} else {
  Resultado_Translacao <- c("Produto apresenta tendência de TRANSLAÇÃO")
}

```

```

# Calcular a discrepância média entre os vértices das linhas de referência e de teste
para correção

# somente vértices da linha com mais vértices e os interpolados da outra linha
}

```

```

### Escala

```

```

Resultado_Escala <- vector()

```

```

Resultado_FE <- vector()

```

```

FE <- vector()

```

```

# Extrair os vértices de início e fim

```

```

for (i in 1:dim(linhas_referencia)[1]) {

```

```

  pts_ref <- st_cast(linhas_referencia_sf[linhas_referencia_sf$ID_Unico == i],
"POINT") # ID_Unico

```

```

  pts_test <- st_cast(linhas_teste_sf[linhas_teste_sf$ID_Unico == i], "POINT") #
ID_Unico

```

```

  start_ref <- pts_ref[1,]

```

```

  end_ref <- pts_ref[nrow(pts_ref),]

```

```

  start_test <- pts_test[1,]

```

```

  end_test <- pts_test[nrow(pts_test),]

```

```

  FE_ref <- sqrt((st_coordinates(start_ref)[1,1] - st_coordinates(end_ref)[1,1])^2 +
                (st_coordinates(start_ref)[1,2] - st_coordinates(end_ref)[1,2])^2)

```

```

  FE_test <- sqrt((st_coordinates(start_test)[1,1] - st_coordinates(end_test)[1,1])^2 +
                (st_coordinates(start_test)[1,2] - st_coordinates(end_test)[1,2])^2)

```

```

  FE[i] <- FE_ref/FE_test

```

```

  if(FE[i] >= 0.9 && FE[i] <= 1.1){

```

```

Resultado_FE[i] <- c("Não apresenta tendência com efeito de escala")
} else {
Resultado_FE[i] <- c("Apresenta tendência com efeito de escala")
}
}

if(all(Resultado_FE == "Não apresenta tendência com efeito de escala")){
Resultado_Escala <- c("Produto sem tendência de ESCALA")
} else {
MP_num <- vector()
MP_denom <- vector()
for(i in 1:dim(linhas_referencia)[1]){
comprim_ref <- st_length(linhas_referencia_sf[linhas_referencia_sf$ID_Unico ==
i,]) # ID_Unico
MP_num[i] <- comprim_ref*FE[i]
MP_denom[i] <- comprim_ref
}
MP <- sum(MP_num)/sum(MP_denom)
if(MP >= 0.9 && MP <= 1.1){
Resultado_Escala <- c("Produto sem tendência de ESCALA")
} else {
Resultado_Escala <- c("Produto apresenta tendência de ESCALA")
}
}
}

```

### Rotação

```

ID_rot <- vector()
Az_rot_ref <- vector()
Az_rot_test <- vector()
Theta_rot <- vector()

# Extrair os vértices de início e fim
for (i in 1:dim(linhas_referencia)[1]) {
  pts_ref <- st_cast(linhas_referencia_sf[linhas_referencia_sf$ID_Unico == i],
"POINT") # ID_Unico
  pts_test <- st_cast(linhas_teste_sf[linhas_teste_sf$ID_Unico == i], "POINT") #
ID_Unico
  start_ref <- pts_ref[1,]
  end_ref <- pts_ref[nrow(pts_ref),]
  start_test <- pts_test[1,]
  end_test <- pts_test[nrow(pts_test),]

  ID_rot[i] <- i # ID_rot == ID_Unico
  Az_rot_ref[i] <- ngeo::st_azimuth(start_ref, end_ref)
  Az_rot_test[i] <- ngeo::st_azimuth(start_test, end_test)
  Theta_rot[i] <- max(Az_rot_ref[i], Az_rot_test[i]) - min(Az_rot_ref[i], Az_rot_test[i])
  if(Theta_rot[i] > 180){
    Theta_rot[i] <- 360 - Theta_rot[i]
  }
}

Rotacao <- data.frame(ID_rot, Az_rot_ref, Az_rot_test, Theta_rot)

```

```

MP_num_rot <- vector()
MP_denom_rot <- vector()

for (i in 1:dim(linhas_referencia)[1]) {
  MP_num_rot[i] <- sin(pracma::deg2rad(Rotacao[Rotacao$ID_rot ==
i,]$Theta_rot))*st_length(linhas_referencia_sf[linhas_referencia_sf$ID_Unico == i,]) #
ID_Unico
  MP_denom_rot[i] <- st_length(linhas_referencia_sf[linhas_referencia_sf$ID_Unico
== i,]) # ID_Unico
}

MP_rot <- sum(MP_num_rot)/sum(MP_denom_rot)

if(MP_rot >= -0.1 && MP_rot <= 0.1){
  Resultado_Rotacao <- c("Produto sem tendência de ROTAÇÃO")
} else {
  Resultado_Rotacao <- c("Produto apresenta tendência de ROTAÇÃO")
}

```