

UNIVERSIDADE FEDERAL DE VIÇOSA

**Estimativa de velocidades a partir de séries temporais posicionais para
atualização temporal de coordenadas: avaliação integrada de métodos e
tratamento de dados**

Krisley Xavier Soares de Freitas
Doctor Scientiae

**VIÇOSA - MINAS GERAIS
2025**

KRISLEY XAVIER SOARES DE FREITAS

**Estimativa de velocidades a partir de séries temporais posicionais para
atualização temporal de coordenadas: avaliação integrada de métodos e
tratamento de dados**

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: William Rodrigo Dal Poz

Coorientador: Afonso de Paula dos Santos

**VIÇOSA - MINAS GERAIS
2025**

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

T

F866e Freitas, Krisley Xavier Soares de, 1991-
2025 Estimativa de velocidades a partir de séries temporais
posicionais para atualização temporal de coordenadas: avaliação
integrada de métodos e tratamento de dados / Krisley Xavier
Soares de Freitas. – Viçosa, MG, 2025.

1 tese eletrônica (149 f.): il. (algumas color.).

Inclui apêndices.

Orientador: William Rodrigo Dal Poz.

Tese (doutorado) - Universidade Federal de Viçosa,
Departamento de Engenharia Civil, 2025.

Referências bibliográficas: f. 132-145.

DOI: <https://doi.org/10.47328/ufvbbt.2025.788>

Modo de acesso: World Wide Web.

1. Coordenadas geográficas - Medição - Métodos
estatísticos. 2. Sistema de Posicionamento Global. 3. Análise de
séries temporais. 4. Outliers (Estatística). I. Dal Poz, William
Rodrigo, 1975-. II. Universidade Federal de Viçosa.
Departamento de Engenharia Civil. Programa de Pós-Graduação
em Engenharia Civil. III. Título.

CDD 22. ed. 526.6402870727

KRISLEY XAVIER SOARES DE FREITAS

**Estimativa de velocidades a partir de séries temporais posicionais para
atualização temporal de coordenadas: avaliação integrada de métodos e
tratamento de dados**

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: 19 de novembro de 2025.

Assentimento:

Krisley Xavier Soares de Freitas
Autor

William Rodrigo Dal Poz
Orientador

Essa tese foi assinada digitalmente pelo autor em 27/11/2025 às 19:54:22 e pelo orientador em 28/11/2025 às 07:11:07. As assinaturas têm validade legal, conforme o disposto na Medida Provisória 2.200-2/2001 e na Resolução nº 37/2012 do CONARQ. Para conferir a autenticidade, acesse <https://siadoc.ufv.br/validar-documento>. No campo 'Código de registro', informe o código **G5WS.E9NC.5TCR** e clique no botão 'Validar documento'.

AGRADECIMENTOS

Aos meus pais Jane e José pelo suporte e incentivo que me possibilitaram perseguir o caminho acadêmico.

Ao amigo Lécio Alves Nascimento pelas inúmeras discussões e conselhos.

Ao professor William Rodrigo Dal Poz pela orientação e apoio.

Aos professores da banca de projeto e exame de qualificação Afonso de Paula dos Santos, Haroldo Antônio Marques, Ítalo Oliveira Ferreira e Júlio César de Oliveira pelas correções e sugestões que contribuíram significativamente para o aprimoramento desta pesquisa.

À Universidade Federal de Viçosa – UFV, em especial ao Programa de Pós-Graduação em Engenharia Civil, pelos conhecimentos e ensinamentos ao longo da graduação, mestrado e doutorado.

Ao Nevada Geodetic Laboratory pela disponibilização de séries temporais posicionais de forma gratuita e sem registro, que contribui de forma incomensurável às pesquisas científicas.

A todos que direta ou indiretamente contribuíram para a realização desta pesquisa.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudos.

Este trabalho foi realizado com o apoio das seguintes agências de pesquisa brasileiras: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

RESUMO

FREITAS, Krisley Xavier Soares de, D.Sc., Universidade Federal de Viçosa, novembro de 2025. **Estimativa de velocidades a partir de séries temporais posicionais para atualização temporal de coordenadas: avaliação integrada de métodos e tratamento de dados.** Orientador: William Rodrigo Dal Poz. Coorientador: Afonso de Paula dos Santos.

Esta pesquisa teve como objetivo analisar a influência que outliers presentes em séries temporais posicionais exercem nas velocidades estimadas a partir destas séries, assim como realizar a otimização do processo de estimativa de velocidades, indo desde a obtenção das séries com um software desenvolvido (GPTS Downloader) até a interpolação espacial das velocidades para pontos de interesse. A partir de séries de 20 anos de dados de 113 estações dos Estados Unidos, estimou-se a velocidade dessas estações, atualizaram-se suas coordenadas e calcularam-se as discrepâncias planimétricas em relação às coordenadas de referência. Empregando séries temporais de dados reais, que refletem a complexidade e a variação natural intrínseca da Terra, verificou-se que a remoção de outliers das séries não apresentou impacto significativo nas discrepâncias planimétricas finais. Ou seja, as velocidades não se mostraram sensíveis aos outliers das séries das quais são estimadas. Analisando técnicas de estimativa de velocidades, verificou-se que os métodos baseados nos mínimos quadrados são intercambiáveis, uma vez que seus resultados mostraram diferenças desprezíveis para esta finalidade. As velocidades estimadas com a regressão linear simples e o MIDAS (Median Interannual Difference Adjusted for Skewness) também foram similares para a maioria das estações, mas o MIDAS apresentou resultados significativamente piores em algumas séries com saltos. No que se refere às multivelocidades, não foi possível prever quando ou qual multivelocidade de período uniforme apresentará melhores resultados que a monovelocidade. No entanto, para séries com saltos ou mudanças de tendência, a multivelocidade com períodos distintos permite uma melhor adequação de cada segmento ao comportamento das séries, reduzindo a discrepância após a atualização das coordenadas. Em relação à interpolação, por apresentar os resultados mais consistentes entre os métodos analisados, a interpolação por Funções de Bases Radiais se mostrou o método mais confiável para ser empregado em velocidades estimadas a partir de séries temporais. Dessa forma, mesmo analisando regiões de alta atividade sísmica, foi possível atualizar coordenadas por 20 anos com baixas discrepâncias planimétricas, apresentando resultados superiores aos dos modelos de velocidades convencionais, mesmo após interpolação.

Palavras-chave: geodésia; GPS; outlier; velocidades geodésicas; série temporal posicional

ABSTRACT

FREITAS, Krisley Xavier Soares de, D.Sc., Universidade Federal de Viçosa, November, 2025. **Velocity estimation from positional time series for coordinate updating: an integrated assessment of methods and data processing.** Adviser: William Rodrigo Dal Poz. Co-adviser: Afonso de Paula dos Santos.

This research aimed to analyze the influence of outliers present in positional time series on the velocities estimated from these series, as well as to optimize the velocity estimation process, from obtaining the series with a developed software (GPTS Downloader) to the spatial interpolation of velocities for points of interest. Based on 20 years of data from 113 stations in the United States, the velocities of these stations were estimated, their coordinates were updated, and the planimetric discrepancies in relation to the reference coordinates were calculated. Using real data time series, which reflect the complexity and intrinsic variation of nature, it was found that removing outliers from the series did not have a significant impact on the final planimetric discrepancies. In other words, the velocities were not sensitive to the outliers of the series from which they were estimated. Analyzing velocity estimation techniques, it was found that least squares-based methods are interchangeable, as their results showed negligible differences for this purpose. The velocities estimated with simple linear regression and MIDAS (Median Interannual Difference Adjusted for Skewness) were also similar for most stations, but MIDAS showed significantly worse results in some series with jumps. Regarding multivelocities, it was not possible to predict when or which multivelocidade with a homogeneous period would yield better results than the monovelocity. However, for series with jumps or trend changes, the multivelocities with distinct periods allows for better adjustment of each segment to the series' behavior, reducing discrepancies after coordinate updates. As for interpolation, due to presenting the most consistent results among the methods analyzed, interpolation using Radial Basis Functions proved to be the most reliable method to be employed for estimating velocities from time series. In this way, even when analyzing regions of high seismic activity, it was possible to update coordinates over 20 years with low planimetric discrepancies, showing results superior to those of conventional velocity models, even after interpolation.

Keywords: geodesy; GPS; outlier; geodetic velocities; positional time series

ÍNDICE DE ILUSTRAÇÕES

| | |
|--|----|
| Figura 1 - Quantidade de estações e os respectivos períodos das séries temporais posicionais disponibilizadas pelo NGL. | 23 |
| Figura 2 - Disposição e período de dados das 22.960 estações englobadas pelo NGL em outubro de 2025. | 24 |
| Figura 3 – Identificação de <i>outliers</i> com <i>Z-Score</i> em uma série temporal de coordenadas diárias (componente X). | 31 |
| Figura 4 - Identificação de <i>outliers</i> com KNN em uma série temporal posicional de dados diários (componente X). | 34 |
| Figura 5 - Identificação de <i>outliers</i> com IF em uma série temporal posicional de dados diários (componente X). | 35 |
| Figura 6 - Identificação de <i>outliers</i> com DBSCAN em uma série temporal posicional de dados diários (componente X). | 36 |
| Figura 7 - Disposição das 113 estações no território dos Estados Unidos da América. | 44 |
| Figura 8 - Fluxograma da metodologia. | 45 |
| Figura 9 –Interface do GPTS Downloader. | 46 |
| Figura 10 - Recorte da interface do GPTS Downloader, utilização do filtro do Brasil e classificação das estações por maior período de dados. | 47 |
| Figura 11 - Fluxograma da obtenção da menor discrepância com os métodos de identificação e remoção de <i>outliers</i> | 50 |
| Figura 12 - Multivelocidades de 4 anos (setas vermelhas) e monovelocidade de 20 anos (seta verde) para a série de coordenada Z da estação MAWY. | 54 |
| Figura 13 - Fluxograma do processo de estimativa das multivelocidades. | 55 |
| Figura 14 - Atualização das coordenadas referentes ao ano de 2019 para 2000,0, utilizando multivelocidades. | 56 |
| Figura 15 - Localização das estações selecionadas para serem consideradas como desconhecidas. | 58 |
| Figura 16 - Frequência de melhor discrepância por método para as 113 estações. | 59 |
| Figura 17 - Séries temporais posicionais com dados preenchidos com os métodos <i>spline</i> , <i>pad</i> , <i>cubicspline</i> e <i>akima</i> | 60 |
| Figura 18 - Discrepância média das 113 estações para cada método. | 61 |

| | |
|--|----|
| Figura 19 - Discrepância planimétrica média das 113 estações por método de estimativa de velocidade..... | 64 |
| Figura 20 - Incertezas médias das velocidades das 113 estações por método de estimativa. .. | 65 |
| Figura 21 - Discrepância planimétrica média para cada estação com o ME, MIDAS e regressão linear. | 66 |
| Figura 22 - Séries temporais posicionais das estações CAND, LAND, MIDA e POMM. | 67 |
| Figura 23 - Discrepância planimétrica média das 113 estações obtidas com as velocidades estimadas a partir das séries temporais posicionais e com 18 modelos de velocidades..... | 69 |
| Figura 24 - Discrepância planimétrica média por método de interpolação para cada estação. | 72 |
| Figura 25 - Discrepância planimétrica média das seis estações consideradas desconhecidas: velocidades interpoladas e 18 modelos de velocidades..... | 74 |
| Figura 26 - Discrepância Sem Remoção de Outliers (DSRO) e discrepâncias após aplicação de média móvel para as estações AGMT, AVRY, CBHS e S300. | 76 |
| Figura 27 - Frequência de tamanho de janela que proporcionou a menor discrepância nas 113 estações. | 76 |
| Figura 28 - <i>Outliers</i> identificados com <i>boxplot</i> ajustado para a estação S300..... | 79 |
| Figura 29 - <i>Outliers</i> identificados com o <i>boxplot</i> ajustado nas estações BRAN, LMUT e RDOM. | 80 |
| Figura 30 - Frequência do <i>threshold</i> que proporcionou a menor discrepância para as 113 estações, após aplicação do <i>boxplot</i> ajustado. | 81 |
| Figura 31 - Discrepâncias para cada valor de <i>threshold</i> nas estações ECHO e LBC2..... | 82 |
| Figura 32 – <i>Outliers</i> identificados nas séries da estação GARL. | 84 |
| Figura 33 - Discrepâncias para os valores de <i>Z-Score</i> das estações CTMS e WMAP..... | 85 |
| Figura 34 - Frequência do <i>Z-Score</i> que proporcionou a menor discrepância para as 113 estações, após aplicação do <i>Z-Score</i> | 86 |
| Figura 35 – Série de coordenada X da estação KOKB, com <i>outliers</i> identificados com <i>Z-Score</i> (esquerda) e <i>Z-Score</i> modificado (direita)..... | 88 |
| Figura 36 - Frequência de <i>threshold</i> que proporcionou a menor discrepância após aplicação do <i>Z-Score</i> modificado. | 89 |
| Figura 37 – <i>Outliers</i> detectados nas séries das estações CJTR, DHLG e KNOL com CBP-HI. | 91 |
| Figura 38 - Valores de <i>cp</i> e suas discrepâncias correspondentes. | 92 |

| | |
|--|-----|
| Figura 39 - Frequência de <i>cp</i> que proporcionou a menor discrepância após aplicação BCP-HI. | 93 |
| Figura 40 - <i>Outliers</i> identificados com KNN. | 95 |
| Figura 41 - Discrepâncias com a variação da contaminação e do número de vizinhos para a estação MIDA. | 96 |
| Figura 42 - Frequência da contaminação que proporcionou a menor discrepância após aplicação do KNN. | 97 |
| Figura 43 - Frequência dos números de vizinhos que proporcionaram as menores discrepâncias após aplicação do KNN. | 97 |
| Figura 44 - <i>Outliers</i> identificados com IF nas estações MIDA e POMM. | 99 |
| Figura 45 - Relação da discrepância com a variação da contaminação com o IF nas estações CAND e OPCP. | 100 |
| Figura 46 - Frequência da contaminação que proporcionou a menor discrepância após aplicação do IF. | 101 |
| Figura 47 - <i>Outliers</i> identificados com DBSCAN nas séries das estações BURN e OPCP. | 104 |
| Figura 48 – Séries da estação OAES com <i>outliers</i> identificados com DBSCAN e após remoção dos <i>outliers</i> | 105 |
| Figura 49 - Séries da estação ORES com <i>outliers</i> identificados com DBSCAN e após remoção dos <i>outliers</i> | 106 |
| Figura 50 - Variação da discrepância média em relação ao valor do parâmetro “número mínimo de pontos da amostra” nas estações CAST e COON. | 107 |
| Figura 51 - Frequência do número mínimo de pontos da amostra que proporcionou a menor discrepância após aplicação do DBSCAN. | 108 |
| Figura 52 - Frequência de ϵ que proporcionou a menor discrepância após aplicação do DBSCAN. | 108 |
| Figura 53 - Frequência em que cada método obteve a menor discrepância entre as 113 estações. | 110 |
| Figura 54 - Número de estações em que a discrepância aumentou após a remoção de <i>outliers</i> e número de estações em que houve uma redução da discrepância superior à 0,001 mm. | 111 |
| Figura 55 - Redução e aumento médio da discrepância planimétrica após remoção de <i>outliers</i> | 111 |
| Figura 56 - Séries da estação LMUT com <i>outliers</i> identificados com DBSCAN. | 113 |

| | |
|---|-----|
| Figura 57 - Incerteza média das velocidades estimadas a partir das séries originais e das séries após remoção de <i>outliers</i> | 114 |
| Figura 58 - O R^2 ajustado médio das 113 estações para as coordenadas X, Y e Z. | 115 |
| Figura 59 - Comparação do número de estações com redução de discrepância maior que 0,001 m após a remoção de <i>outliers</i> em séries sem tendência e sem tendência e sazonalidade. | 116 |
| Figura 60 - Discrepância planimétrica média para as estações AGMT, AVRY, LDES e MWTP para cada velocidade..... | 117 |
| Figura 61 - Frequência da velocidade que resultou na melhor discrepância planimétrica calculada para as 113 estações..... | 118 |
| Figura 62 – Frequência de menor discrepância planimétrica na comparação de cada multivelocidade isoladamente com a monovelocidade (rosa) para as 113 estações. | 119 |
| Figura 63 - Discrepância planimétrica média das 113 estações para cada velocidade..... | 120 |
| Figura 64 - Segmentação das séries das estações BMHL, DHLG, MAWY e NAIU. | 122 |
| Figura 65 - Segmentação das séries das estações NLIB, OPBL, OPCP e OPCX..... | 123 |
| Figura 66 - Segmentação das séries das estações OPRD e SLMS. | 124 |
| Figura 67 - Discrepâncias médias calculadas após atualização de coordenadas com mono e multivelocidades. | 125 |
| Figura 68 - Incertezas das multivelocidades das estações MAWY e OPCP..... | 126 |
| Figura 69 - Incertezas das coordenadas atualizadas com multivelocidade das estações BMHL, DHLG, MAWY e NAIU. | 127 |
| Figura 70 - Incertezas das coordenadas atualizadas com multivelocidade das estações NLIB, OPBL, OPCP e OPCX. | 128 |
| Figura 71 - Incertezas das coordenadas atualizadas com multivelocidade das estações OPDR e SLMS..... | 129 |

ÍNDICE DE TABELAS

| | |
|---|-----|
| Tabela 1 - Velocidades estimadas com os diferentes métodos..... | 63 |
| Tabela 2 - Discrepâncias antes e depois da aplicação de média móvel simples. | 77 |
| Tabela 3 - Discrepâncias antes e após remoção de <i>outliers</i> com <i>boxplot</i> ajustado. | 78 |
| Tabela 4 - Discrepâncias antes e após eliminação de <i>outliers</i> com <i>Z-Score</i> | 83 |
| Tabela 5 - Discrepâncias antes e após eliminação de <i>outliers</i> com <i>Z-Score</i> modificado..... | 87 |
| Tabela 6 - Discrepâncias antes e após eliminação de <i>outliers</i> com BCP-HI..... | 90 |
| Tabela 7 - Discrepâncias antes e após eliminação de <i>outliers</i> com KNN..... | 94 |
| Tabela 8 - Discrepâncias antes e após eliminação de <i>outliers</i> com IF..... | 98 |
| Tabela 9 - Discrepâncias antes e após eliminação de <i>outliers</i> com DBSCAN..... | 102 |
| Tabela 10 - Estações que apresentaram redução de discrepância acima de 0,001 m e o respectivo método..... | 112 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|----------|--|
| CP | <i>Change Points</i> |
| CSRS-PPP | <i>Canadian Spatial Reference System – PPP</i> |
| DBSCAN | <i>Density-Based Spatial Clustering of Applications with Noise</i> |
| DSRO | Discrepância sem remoção de <i>outliers</i> |
| GNSS | <i>Global Navigation Satellite Systems</i> |
| GPS | <i>Global Positioning System</i> |
| IDW | <i>Inverse distance weighted</i> |
| IGS | <i>International GNSS Service</i> |
| ITRF | <i>International Terrestrial Reference Frame</i> |
| KO | Krigagem Ordinária |
| MAD | Desvio Absoluto da Mediana |
| MIDAS | <i>Median Interannual Difference Adjusted for Skewness</i> |
| ME | <i>Minimum Entropy</i> |
| NGL | <i>Nevada Geodetic Laboratory</i> |
| RBF | <i>Radial Basis Function interpolation</i> |
| PPP | Posicionamento por Ponto Preciso |
| PSD | <i>Post-Seismic Deformation</i> |
| RINEX | <i>Receiver Independent Exchange Format</i> |
| SIRGAS | Sistema de Referência Geocêntrico para as Américas |
| UNAVCO | <i>University NAVstar COnsortiu</i> |
| WLS | <i>Weighted least squares</i> |

SUMÁRIO

| | | |
|-------|--|----|
| 1 | INTRODUÇÃO | 15 |
| 1.1 | Hipóteses | 20 |
| 2 | OBJETIVOS..... | 21 |
| 2.1 | Objetivo geral | 21 |
| 2.2 | Objetivos específicos..... | 21 |
| 3 | JUSTIFICATIVA..... | 21 |
| 4 | REVISÃO DE LITERATURA | 22 |
| 4.1 | Séries temporais posicionais do NGL | 22 |
| 4.2 | Imputação de dados ausentes..... | 25 |
| 4.3 | Estimativa de velocidades | 26 |
| 4.4 | Identificação e remoção de <i>outliers</i> | 30 |
| 4.5 | Interpolação de velocidades..... | 38 |
| 4.6 | Atualização temporal de coordenadas | 41 |
| 4.7 | Cálculo da discrepância planimétrica | 42 |
| 5 | MATERIAIS E MÉTODOS | 43 |
| 5.1 | Materiais | 43 |
| 5.2 | Metodologia..... | 44 |
| 5.2.1 | GPTS Downloader..... | 45 |
| 5.2.2 | Imputação de dados ausentes..... | 48 |
| 5.2.3 | Identificação e remoção de outliers | 49 |
| 5.2.4 | Estimativa de velocidades | 51 |
| 5.2.5 | Multivelocidade | 53 |
| 5.2.6 | Coordenadas de referência..... | 56 |
| 5.2.7 | Interpolação de velocidades..... | 57 |
| 6 | RESULTADOS..... | 58 |
| 6.1 | Imputação de dados ausentes..... | 58 |

| | | |
|--------|--|-----|
| 6.2 | Estimativa de velocidades | 61 |
| 6.3 | Interpolação de velocidades..... | 70 |
| 6.4 | Identificação e remoção de <i>outliers</i> | 75 |
| 6.4.1 | Média móvel simples..... | 75 |
| 6.4.2 | Boxplot Ajustado..... | 78 |
| 6.4.3 | Z-Score | 82 |
| 6.4.4 | Z-Score Modificado..... | 86 |
| 6.4.5 | BCP-HI..... | 89 |
| 6.4.6 | KNN | 93 |
| 6.4.7 | Isolation Forest | 98 |
| 6.4.8 | DBSCAN - Density-Based Spatial Clustering of Applications with Noise | 101 |
| 6.4.9 | Comparações dos métodos | 109 |
| 6.4.10 | Remoção de tendência e sazonalidade..... | 115 |
| 6.5 | Multivelocidade..... | 116 |
| 7 | CONCLUSÕES..... | 129 |
| | REFERÊNCIAS | 132 |
| | APÊNDICES | 146 |

1 INTRODUÇÃO

As técnicas geodésicas baseadas em satélites fornecem uma ferramenta excepcional para monitorar e explorar uma ampla gama de processos geofísicos no sistema terrestre, dentre eles: terremotos (Park; Grejner-Brzezinska; Von Frese, 2014; Pavani *et al.*, 2016), derretimento e acumulação de neve (Najibi; Jin; Wu, 2015), deformações crustais (Khazaradze; Echeverria; Asensio, 2014; Tatevian *et al.*, 2014, Danezis; Chatzinikos; Kotsakis, 2020), marés terrestres (Pahlevi *et al.*, 2017) e carga hidrológica (Nascimento *et al.*, 2021). Tais processos provocam deslocamentos na crosta e geram variações espaço-temporais de diferentes magnitudes na posição de pontos na superfície terrestre (Sapucci; Monico, 2001; Sánchez; Drewes, 2016).

De acordo com SIRGAS (2025a), é possível obter as velocidades desses pontos através de modelos de deformações ou velocidades, ou estimá-las através de séries temporais posicionais oriundas de levantamentos GNSS (*Global Navigation Satellite Systems*), realizados de forma contínua por pelo menos dois anos.

Embora os programas de processamento de observações GNSS venham proporcionado resultados mais confiáveis, as séries temporais posicionais ainda são afetadas por três tipos de fatores: o primeiro tipo está associado ao processamento de dados brutos, como correções do centro de fase da antena, órbita dos satélites ou atraso atmosférico; o segundo diz respeito ao ambiente da estação GNSS, como a estabilidade do monumento e multicaminho; o terceiro relaciona-se à análise de pós-processamento das observações, como a definição do referencial, sinais periódicos e deslocamentos de posição devido a troca de equipamentos, terremotos ou fontes indefinidas (Cannavò *et al.*, 2010; Masson; Mazzotti; Vernant, 2019).

O efeito desses fatores nas séries pode influenciar consideravelmente as velocidades estimadas a partir dessas séries, devido a, por exemplo, ocorrência de saltos, ausência de dados e presença de *outliers*.

A ausência de dados é um problema comum em pesquisas científicas e podem possuir diversas causas (Newsom, 2025). Para o caso das séries temporais posicionais, além dos fatores supracitados, destaca-se falhas ou defeitos nos receptores GNSS, falta de energia elétrica e problemas com servidor ou com o armazenamento de dados (Liu; Li, 2022).

Esses dados ausentes em séries temporais geram três preocupações principais: i) a redução do tamanho da amostra - à medida que a quantidade de dados ausentes aumenta, o tamanho da amostra diminui. Com menos dados disponíveis, torna-se mais difícil detectar efeitos ou tendências nas séries, prejudicando a qualidade das análises (Neto e Cascudo, 2024;

Newsom, 2025); ii) viés e distorção nos resultados - os dados ausentes podem introduzir viés nas estimativas dos coeficientes de regressão, dos termos de erro e dos testes de significância. Isso ocorre principalmente quando os valores ausentes não são completamente aleatórios (isto é, se a ausência dos dados estiver relacionada com as próprias variáveis estudadas), fazendo com que as estimativas possam ser distorcidas ou enganosas (Newsom, 2025); e, iii) impacto na inferência estatística - os dados ausentes também afetam os testes de hipótese e a inferência estatística de maneira geral. Eles podem fazer com que os erros-padrões fiquem inflacionados ou subestimados, o que altera tanto os intervalos de confiança como os valores de p , levando a conclusões equivocadas sobre a significância dos resultados (Thomas; Rajabi, 2021).

A imputação de dados ausentes pode ser realizada através de diversas abordagens, cujo método específico a ser empregado para obter os melhores resultados depende das características das séries temporais analisadas (Neto e Cascudo, 2024). Métodos como a média móvel (Box; Jenkins; Reinsel, 2015), *spline* (De Boor, 2001) e repetição dos valores disponíveis mais próximos (Mallinckrodt *et al.*, 2008) são empregados em diversos tipos de dados. Para as séries temporais posicionais, geralmente emprega-se a interpolação linear para realizar a imputação, como realizado por Moritz *et al.* (2015), Klos, Bos e Bogusz (2018) e Freitas, Dal Poz e Nascimento (2022).

Além dos dados ausentes, outro fator que pode afetar a estimativa de velocidades é a presença de *outliers* nas séries temporais. Segundo Hawkins (1980, p. 1, tradução própria), *outlier* é “uma observação que desvia tanto das outras observações a ponto de levantar suspeita de que ela foi gerada por um mecanismo diferente”. A detecção de *outliers* é um problema amplamente tratado na literatura, embora ainda não esteja totalmente desenvolvido para o tipo específico de séries temporais posicionais devido à natureza dinâmica e as variadas aplicações no mundo real desses dados (Cannavò *et al.*, 2010; Hassan; Barakat; Rezk, 2022). Segundo Borradaile (2003), dados de ciências da Terra (*Earth Science Data*) refletem a complexidade e a variação natural intrínseca da natureza, não sendo possível ter controle sobre esses dados. Consequentemente, as séries temporais posicionais refletem essa complexidade.

Como destaca Shen e Yang (2022), o processo de detecção de *outliers* tem sido extensivamente aplicado em diversos campos científicos e tecnológicos, como dados biológicos (Shetta; Niranjana, 2020; Luo; Gao; Halabi, 2022), sistemas industriais (Das; Adepu; Zhou, 2020), transações de cartão de crédito (John; Naaz, 2019; Lokanan, 2022), texto (Gorokhov; Petrovskiy; Mashechkin, 2017; Park, 2022) e *streaming data* (Hassan; Barakat; Rezk, 2022).

Em se tratando de séries temporais posicionais, a detecção de *outliers* vem sendo empregada, entre outras aplicações, para melhoria do processo da estimativa do atraso instrumental de receptores GNSS (Amaral; Klein; De Oliveira Junior, 2022), detecção de descontinuidades em séries (Crocetti; Schartner; Soja, 2021) e melhoria no rastreamento de celulares via GPS (*Global Positioning System*) (Pereira; Sofia, 2022). Para a finalidade de pré-processamento dessas séries, ainda não há um consenso sobre quais técnicas ou metodologia utilizar para a detecção de *outliers*, sendo empregadas desde metodologias mais simples, como as baseadas na distribuição dos dados, até métodos mais elaborados, com utilização de aprendizado de máquina e redes neurais.

Khazaradze, Echeverria e Asensio (2014) detectaram *outliers* seguindo dois critérios: se os erros de posição fossem superiores a 20 mm e caso a própria posição estivesse deslocada mais de 10 mm em relação à linha de tendência ajustada à série.

Nascimento *et al.* (2017) e Freitas, Nascimento e Dal Poz (2022) utilizaram um filtro de médias móveis simples de ordem 12 visando reduzir oscilações dos ruídos e suavizar as curvas da série. Lima (2015) utiliza filtros de média móvel simples de ordem 6 em séries temporais horárias de dados GNSS para monitoramento de deslocamento em estruturas. Le Thi *et al.* (2021) que usaram filtros de ordem 30, salientam que a média móvel é sensível à *outliers* e que são adequadas para suavizar séries temporais.

Li *et al.* (2000) empregaram o método da Média Móvel Autoregressiva (ARMA) em séries temporais posicionais GPS do Japão, objetivando remover ruído branco (valores aleatórios independentes com distribuição fixa e que não apresentam tendência, sazonalidade ou previsibilidade) e detectar descontinuidades e *outliers*, com a finalidade de gerar campos de velocidades.

Ji e Shen (2020) propõem uma abordagem que primeiro recupera os sinais temporais das séries temporais de posição GNSS usando análise de *wavelets*, e depois detecta *outliers* na série temporal residual de posição usando o intervalo interquartil.

Klos *et al.* (2015) empregam tanto o método 3σ quanto o 5σ . Kierulf *et al.* (2021) e Savchyn, Brusak e Tretyak (2022) utilizam o 3σ . Crocetti, Schartner e Soja (2021) utilizam 5σ .

Qianqian e Qingming (2013) propõe um método Bayesiano para detectar grupos de *outliers* baseado no amostrador de Gibbs. Pehlivan (2024) combina a análise do Ponto de Mudança Bayesiano com o identificador de Hampel.

Nguyen e Tran (2023) avaliaram a performance de métodos de aprendizado de máquina - *Isolation Forest* (IF), *One-Class Support Vector Machines* (O-C SVM) e *Local Outlier Factor*

(LOF) – na detecção de *outliers* em séries temporais GNSS sintéticas. Kiani (2020) desenvolveu um algoritmo de aprendizado de máquina para prever séries temporais de posição GNSS, detectar *outliers* e anomalias e prever terremotos.

Em diversas pesquisas é prática comum utilizar séries sintéticas ou imputar *outliers* artificiais nas séries como forma de controle para avaliar a performance da metodologia testada: Tran *et al.* (2016) analisam os métodos de detecção de *outliers* testados através de identificação visuais e de métricas estatísticas; Kiani (2020) inseriu 10 mil *outliers* simulados (com magnitude entre 2 e 5 cm) em 2 mil séries sintéticas, com cada série possuindo pelo menos 2 *outliers*; Nguyen e Tran (2023) inseriram *outliers* com amplitude de ± 10 mm a ± 200 mm que compreenderam aproximadamente 15% da série. Entretanto, *outliers* em séries temporais posicionais apresentam magnitude e dispersão que não podem ser simulados com fidelidade devido às suas diversas origens. Desta forma, é importante realizar testes com dados reais e avaliar os métodos de detecção dentro do contexto metodológico em que as séries serão utilizadas. Até o momento da elaboração desta pesquisa, não foi encontrada nenhuma publicação que quantifique o impacto da remoção de *outliers* através da análise de discrepâncias planimétricas (calculadas após a atualização temporal de coordenadas utilizando velocidades estimadas a partir de séries que passaram pela remoção de *outliers*).

Para a estimativa de velocidades a partir de séries temporais posicionais também não há um consenso sobre qual metodologia utilizar. Dentre as mais comuns estão: a inversão de mínimos quadrados do modelo paramétrico (Santamaría-Gómez, 2019; Mazzotti *et al.*, 2020), estimativa de Máxima Verossimilhança das velocidades (Bos *et al.*, 2013; Sakic *et al.*, 2019), *Median Interannual Difference Adjusted for Skewness* - MIDAS (Hammond; Blewitt; Kreemer, 2016; Blewitt *et al.*, 2016; Nistor *et al.*, 2021; Zeng, 2023) e regressão linear simples (Li *et al.*, 2000; Bogusz *et al.*, 2016; Nascimento; Dal Poz; Freitas, 2021; Ramos; Dal Poz; Carvalho, 2021).

Vale destacar que, apesar das pesquisas supracitadas estimarem as velocidades, elas não as aplicam na atualização de coordenadas. Tal etapa possibilita a avaliação das velocidades baseadas nas discrepâncias das coordenadas finais obtidas, assim como mencionado anteriormente no contexto da avaliação da remoção de *outliers*. A prática mais comum é a comparação numérica das velocidades estimadas com as de modelos de velocidades geofísicos e/ou geodésicos. Mahmoud *et al.* (2004) estima um campo de velocidade para a península do Sinai e compara com o NUVEL-1A. Fernandes *et al.* (2007) processa dados de estações da Península Ibérica, França e Norte da África para estimar um campo de velocidades do segmento

Ibero-Magrebino do limite da placa Euroasiática-Somali e compara os resultados com os modelos NUVEL-1A (Demets *et al.*, 1994) e DEOSVel (Fernandes *et al.*, 2003; Fernandes, 2004). Nascimento, Dal Poz e Freitas (2021) estimam velocidades para 9 estações na Bacia Amazônica e comparam com o VEMOS2009 (Drewes; Heidbach, 2012). A desvantagem desta abordagem está no fato de que, ao considerar as velocidades destes modelos como “verdadeiras”, não é possível determinar se os resultados analisados são melhores ou piores quando há divergência de valores.

A estimativa de velocidades através de séries temporais posicionais possibilita a obtenção de velocidade de pontos conhecidos. Entretanto, para a obtenção de velocidades de pontos desconhecidos, deve-se realizar a interpolação dos pontos amostrais. Assim como na detecção de *outliers* e na estimativa de velocidades, diversos métodos são utilizados para realizar a interpolação.

O *Inverse distance weighted* (IDW) é utilizado pelos softwares TREVEL (PROL *et al.*, 2014) e RINEXHub (Freitas; Dal Poz; Nascimento, 2022) para interpolar as velocidades do VEMOS2009. Jiang, Ding e Lan (2024) estimam velocidades verticais de movimentos crustais interpolando estações GNSS de referência na China. Bogusz *et al.* (2014) utilizam o IDW e o vizinho mais próximo para interpolar velocidades de estações permanentes na Polônia.

Outro método comumente utilizado é a krigagem. Liu, Jiang e Wu (2014) interpolam um campo de velocidades de estações GPS para analisar deformações crustais. Almeida *et al.* (2017) geraram mapas de previsão interpolando as velocidades de estações brasileiras. Luna *et al.* (2022) geraram um campo de velocidades para o Equador continental.

Nesse contexto, mostra-se oportuno avaliar os métodos de interpolação para obter as velocidades de pontos desconhecidos utilizando o mesmo critério de avaliação empregado para as etapas metodológicas supracitadas: a discrepância obtida após atualização de coordenadas.

Atualmente, para realizar a atualização de coordenadas de um ponto, inicialmente estima-se a velocidade a partir de uma série temporal posicional, e em seguida atualiza-se as coordenadas para a época de interesse. Tal metodologia leva em consideração que os pontos da superfície terrestre possuem um deslocamento constante no tempo, ou seja, velocidade única. Como destacam Jesus Junior *et al.* (2024), esse modelo linear pode não refletir as complexidades das séries, principalmente devido a padrões periódicos oriundos de eventos geofísicos. No entanto, como proposto por Freitas, Dal Poz e Nascimento (2022a), alternativamente, é possível utilizar multivelocidades. Neste processo, para um mesmo ponto de interesse são atribuídas múltiplas velocidades, com cada uma delas estando associada a um

trecho distinto do período de atualização. Entretanto, destaca-se que tal metodologia ainda não foi extensivamente testada e carece de maior aprofundamento.

Diante do exposto, pode-se notar que ter acesso a séries de dados posicionais de estações GPS estimados com técnicas robustas de posicionamento é essencial para novas aplicações e descobertas científicas. Nesse âmbito, o NGL (*Nevada Geodetic Laboratory*) implementou um sistema que diminui as barreiras do acesso às séries de dados, facilitando assim a realização de inúmeras investigações. O NGL coleta dados brutos de mais de 22.900 estações ao redor do mundo, com algumas contendo um período de dados de mais de 30 anos. No entanto, a disponibilização de uma quantidade tão grande de dados traz consigo o desafio de apresentar aos usuários opções de acesso, visualização espacializada e seleção das estações de interesse de forma simplificada e intuitiva. Atualmente, quem busca estações de um país específico, por exemplo, deve navegar pelo mapa online, identificar o nome das estações de interesse e acessar seus respectivos dados de forma individual, tornando assim o processo moroso e pouco eficiente, especialmente para casos que utilizam muitas estações. Neste sentido, mostra-se pertinente o desenvolvimento de um software destinado à simplificação do processo de seleção das estações de interesse e acesso às séries posicionais disponibilizadas pelo NGL.

1.1 Hipóteses

Diante do exposto, propõem-se as seguintes hipóteses para esta pesquisa:

- i. A aplicação de técnicas para detectar e remover *outliers* em séries temporais posicionais contribui significativamente para a melhoria da estimativa de velocidades, refletindo positivamente nos resultados da atualização temporal de coordenadas;
- ii. A estimativa de velocidades através de séries temporais posicionais e posterior interpolação para pontos de interesse gera resultados melhores na atualização de coordenadas que aqueles obtidos com modelos de velocidades; e,
- iii. Coordenadas atualizadas com multivelocidades apresentam discrepâncias menores em comparação àquelas atualizadas com uma única velocidade.

2 OBJETIVOS

2.1 Objetivo geral

Analisar a influência de *outliers* e otimizar o processo de estimativa de velocidades a partir de séries temporais posicionais para aplicação no processo de atualização temporal de coordenadas.

2.2 Objetivos específicos

Visando atingir o objetivo geral, propõem-se os seguintes objetivos específicos:

- Desenvolver um software para a obtenção de séries temporais posicionais (GPS) de estações situadas ao redor do mundo;
- Avaliar o desempenho de diversos métodos para a imputação de dados ausentes em séries temporais posicionais;
- Avaliar o desempenho de diversos métodos de remoção de *outliers* em séries temporais posicionais, utilizando dados reais;
- Avaliar o desempenho de diversos métodos para a estimativa de velocidades através de séries temporais posicionais em comparação com modelos de velocidades geodésicos e geofísicos;
- Avaliar o desempenho de diversos métodos de interpolação espacial para obter velocidades de pontos desconhecidos; e,
- Avaliar a utilização de multivelocidades na atualização temporal de coordenadas.

3 JUSTIFICATIVA

O GPS é uma ferramenta de fundamental importância para a nova geração das ciências da Terra, pois possibilita investigar e/ou monitorar vulcões, terremotos, tsunamis, alterações no nível do mar, geleiras, aquíferos, deformações crustais, deslocamento de placas tectônicas etc. (Blewitt; Hammond; Kreemer, 2018). Essas investigações geralmente envolvem a análise e aplicação de séries temporais posicionais oriundas de estações geodésicas que realizam monitoramento de forma contínua ao longo de vários anos (He *et al.*, 2017). Tanto as posições como as velocidades dessas estações estão em constante variação devido ao contínuo processo

de deformação da Terra, o que consequentemente afeta também os referenciais, sejam eles globais, como o ITRF (*International Terrestrial Reference Frame*), ou suas realizações intercontinentais, como o SIRGAS (Sánchez; Drewes, 2016).

De forma a garantir a estabilidade de longa duração desses referenciais, é indispensável a utilização de modelos contínuos de deformação crustal (ou de velocidade) para realizar a atualização temporal de coordenadas, como destaca Sánchez e Drewes (2016). Deste modo, a utilização de modelos que não englobem o tempo total da atualização pode reduzir consideravelmente a qualidade dos resultados. No Brasil, por exemplo, onde se adota o SIRGAS2000 (época 2000,4) como o referencial geodésico oficial, isso significa que os levantamentos GNSS realizados atualmente devem ter suas coordenadas atualizadas em mais de duas décadas. Entretanto, não há nenhum modelo disponível que compreenda a totalidade deste período.

Desta forma, é fundamental avaliar os processos metodológicos envolvidos na atualização temporal de coordenadas – imputação de dados ausentes, remoção de *outliers*, estimativa de velocidades, interpolação de velocidades, utilização de multivelocidades –, possibilitando aos profissionais e à comunidade acadêmica selecionar a metodologia que melhor se enquadre a seu objetivo.

4 REVISÃO DE LITERATURA

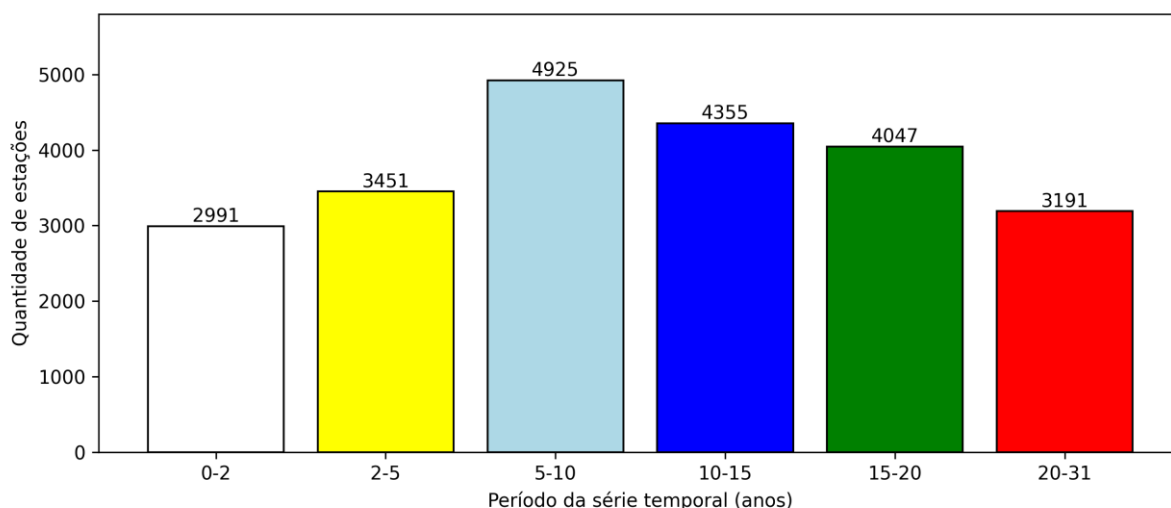
4.1 Séries temporais posicionais do NGL

O NGL busca os dados brutos de observação contínuas – somente da constelação GPS - coletados por estações em mais de 130 repositórios *online* na tentativa de encontrar todos os dados úteis possíveis, reunindo-os em um único local. Tais dados são coletados, gerenciados e distribuídos por centenas de organizações que cooperam através de várias instituições como o IGS (*International GNSS Service*) e UNAVCO (*University NAVstar COnsortiu*). Para lidar com a altíssima massa de dados brutos, o sistema do NGL utiliza novas estratégias de processamento, sistemas automatizados, algoritmos e técnicas de estimativa robustas. Salienta-se que o processamento dos dados brutos é efetuado utilizando o software Gipsy X (Bertiger *et al.*, 2020), cujos parâmetros e produtos aplicados na estratégia de processamento são sumarizados e podem ser consultados em <https://geodesy.unr.edu/gps/ngl.acn.IGS20.txt>. Após o processamento dos dados brutos, são disponibilizadas as séries temporais de dados

posicionais contendo as coordenadas geodésicas e cartesianas com suas respectivas precisões e covariâncias para todas as estações encontradas, com intervalos de observação de 5 minutos e 24h e diferentes latências (tempo entre a coleta dos dados e a disponibilidade para uso) (Blewitt; Hammond; Kreemer, 2018). Destaca-se que todas as séries possuem todas suas coordenadas em um mesmo referencial, o IGS20 (desde agosto de 2025) (NGL, 2025).

Em sua maior parte, as séries abarcam um período entre 5 e 20 anos, como apresentado na Figura 1. Além disso, há um número considerável de séries entre 20 e 30 anos. Nesse sentido, observa-se que as séries posicionais disponibilizadas possibilitam o desenvolvimento de investigações mais abrangentes acerca de efeitos de longa duração, tendências e periodicidades.

Figura 1 - Quantidade de estações e os respectivos períodos das séries temporais posicionais disponibilizadas pelo NGL.



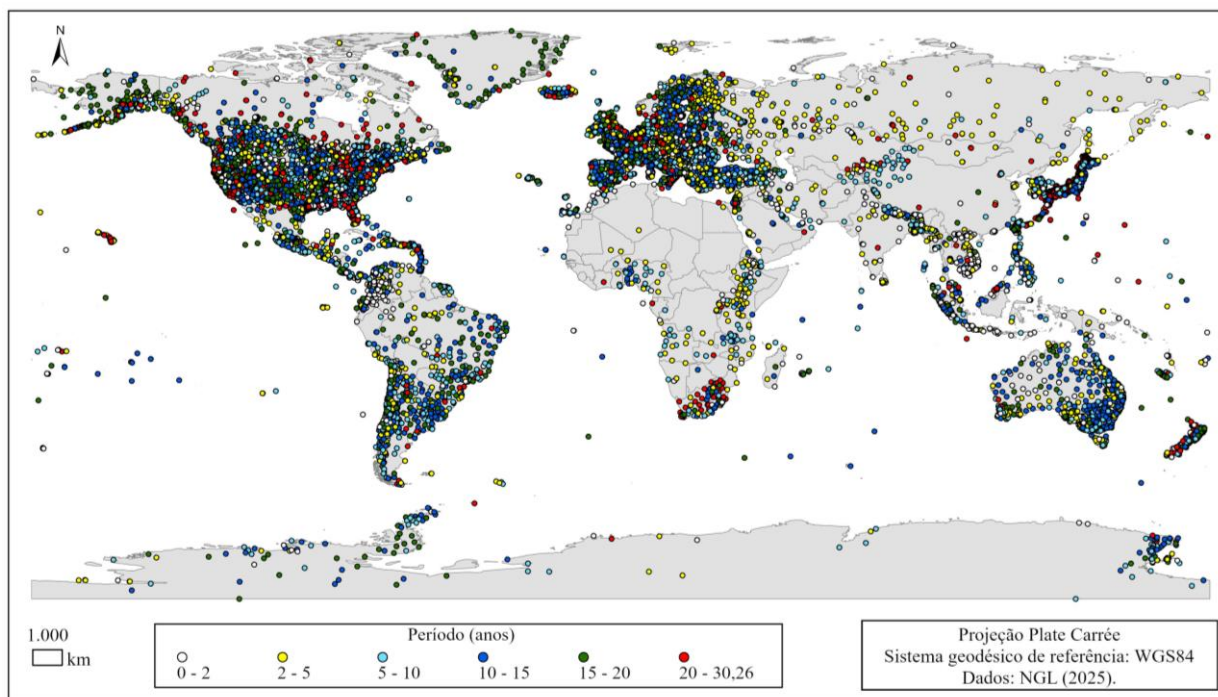
Fonte: elaboração própria.

Nota: dados fornecidos pelo NGL (2025).

Destaca-se que, em outubro de 2025, o NGL atualiza semanalmente as posições diárias de 22.960 estações; e diariamente, as posições de 5 minutos de 10.726 estações (NGL, 2025). Além das coordenadas, as velocidades das estações (estimadas utilizando o MIDAS), também são atualizadas periodicamente.

A disposição das estações, assim como o período das séries posicionais disponibilizadas pelo NGL em 2025 podem ser verificados na Figura 2.

Figura 2 - Disposição e período de dados das 22.960 estações englobadas pelo NGL em outubro de 2025.



Fonte: elaboração própria.

Nota: disponível em alta qualidade em: bit.ly/NGL_Mapa.

O sucesso dessa iniciativa do NGL é dependente da disponibilidade de dados GPS provindos das redes contribuidoras, o que permite a disponibilidade dos dados posicionais de forma aberta e gratuita, possibilitando assim que operadores de redes possam comparar seus resultados com aqueles de outras redes, sendo essa operação benéfica para todas as partes envolvidas. Dessa forma, o NGL e a UNAVCO criaram conjuntamente o projeto *Plug and Play* GPS, fundado pela ACCESS (*Advancing Collaborative Connections for Earth System Science*) da NASA (*National Aeronautics and Space Administration*) (Blewitt; Hammond; Kreemer, 2018). O objetivo deste projeto é reduzir barreiras para pesquisadores menos experientes que desejam participar da comunidade geodésica, como por exemplo, ao instalar uma estação GPS em uma região de cobertura pouco densa, como destacam Baratoux *et al.* (2017).

Os produtos gerados pelo NGL são disponibilizados de forma online e gratuita no endereço <http://geodesy.unr.edu/>, incluindo metadados, listas de estações, gráficos de séries temporais posicionais, tabelas de dados disponíveis, e descrições de novos itens relacionados aos produtos.

4.2 Imputação de dados ausentes

As séries temporais posicionais normalmente apresentam ausência de dados, o que além de causar redução do tamanho da amostra, também pode gerar distorção nos resultados e impactar na inferência estatística (Neto e Cascudo, 2024; Newsom, 2025). Portanto, a imputação desses dados ausentes deve ser realizada.

Diversos métodos de interpolação podem ser empregados nesta imputação, como:

i. *spline* – utiliza polinômios de baixo grau definidos em segmentos adjacentes, cujos pontos de junção são chamados de nós. A suavidade da curva resultante reside nas condições de continuidade impostas nos nós, garantindo que as seções polinomiais se conectem de forma suave em termos de suas derivadas (De Boor, 1978);

ii. PAD (*Piecewise Akima Derivative*) - é uma abordagem de interpolação *piecewise* (por partes) cúbica onde as derivadas nos nós são estimadas localmente, sendo influenciadas pelos pontos vizinhos, e esses valores são utilizados para definir os polinômios interpoladores entre os nós (Akima, 1970, 1978);

iii. *cubicspline* - divide os dados em intervalos e ajusta um polinômio cúbico diferente em cada segmento. Utiliza condições de contorno naturais: a segunda derivada nas extremidades é igual a zero (De Boor, 1978; Press *et al.*, 2007);

iv. *akima* - interpolação por *spline* que evita oscilações excessivas usando derivadas estimadas localmente, que são influenciadas apenas por um pequeno número de pontos vizinhos. Especificamente, para calcular a derivada em um ponto, o método considera as inclinações das retas que conectam os pontos vizinhos; uma média ponderada dessas inclinações é então usada para estimar a derivada no ponto central e os pesos são determinados pela magnitude da diferença entre as inclinações adjacentes, dando um peso menor a inclinações que mudam bruscamente (Akima, 1970, 1978);

v. *cubic* - é um método que usa polinômios de terceiro grau para estimar valores entre pontos de dados, buscando uma representação mais suave da função subjacente em comparação com métodos de ordem inferior (Conte; De Boor, 1980; Burden; Faires, 2010);

vi. PCHIP (*Piecewise Cubic Hermite Interpolating Polynomial*) - é um método de interpolação cúbica *piecewise* que se destaca por sua capacidade de preservar a monotonicidade (ordem consistente) dos dados, evitando oscilações indesejadas e gerando curvas suaves que seguem a tendência dos dados de entrada (Fritsch; Carlson, 1980);

vii. *polynomial* - utiliza um único polinômio para passar por todos os pontos de dados, o que pode apresentar limitações quando o número de pontos é alto, principalmente devido às oscilações (Burden; Faires, 2010);

viii. *nearest* ou interpolação por ponto mais próximo - é um método simples e rápido de interpolação. Ao invés de empregar funções complexas para estimar valores entre os pontos conhecidos, esse método simplesmente atribui a um ponto de interesse o valor do dado conhecido mais próximo (Davis, 1975);

ix. *linear* - é um método simples de interpolação que estima um novo ponto dentro do intervalo de dois pontos de dados conhecidos, assumindo que a função entre esses pontos se comporta como uma linha reta (Chapra; Canale, 2015);

x. *Piecewise polynomial* - é uma técnica que consiste em ajustar uma função aos dados utilizando múltiplos polinômios de grau inferior, cada um aplicado a um subintervalo diferente. A principal vantagem da interpolação polinomial *piecewise* é a sua flexibilidade em lidar com conjuntos de dados grandes e complexos, evitando as oscilações típicas da interpolação polinomial de alto grau. Ao usar polinômios de grau inferior localmente, é possível capturar melhor as variações locais dos dados (Gilat; Subramaniam, 2014);

xi. *slinear* - é um tipo de interpolação polinomial *piecewise* onde cada segmento entre dois pontos de dados consecutivos é interpolado por um polinômio de grau 1, ou seja, uma linha reta (De Boor, 1978);

xii. *time* - realiza uma interpolação linear baseada na distância temporal entre os pontos de dados (Pandas, 2025a);

xiii. *quadratic* - é um método que estima novos pontos de dados dentro do intervalo de três pontos conhecidos, ajustando um polinômio de segundo grau que passa por esses três pontos (Stoer; Bulirsch, 2002); e,

xiv. *from derivatives* - constrói um polinômio interpolador (normalmente cúbicos) utilizando não apenas os valores da função nos pontos conhecidos, mas também suas derivadas (Stoer; Bulirsch, 2002).

4.3 Estimativa de velocidades

Para estimar a velocidade a partir das séries temporais posicionais, pode-se empregar diversos métodos: o MIDAS, como utilizado por Blewitt *et al.* (2016) e Nistor *et al.* (2021); regressão linear simples, como utilizado por Nascimento, Dal Poz e Freitas (2021); *minimum*

entropy, como utilizado por Saleh, Bennett e Williams (2024); e, mínimos quadrados ponderados (*Weighted least squares* - WLS), como utilizado por Santamaría-Gómez (2019).

O modelo de regressão linear simples é dado pela Equação 1 (Draper; Smith, 1998):

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (1)$$

onde y_i é a variável dependente; β_0 é o intercepto de y ; β_1 é a inclinação; x_i é a variável independente; ε_i é o erro aleatório em x ; e, $i = 1, 2, 3, \dots, n$.

β_0 e β_1 são estimados pelos valores b_0 e b_1 , baseados nas amostras de n observações. Assim, a regressão linear amostral é descrita de acordo com a Equação 2, onde \hat{y}_i indica o valor estimado da variável dependente, desconsiderando o erro aleatório (Draper; Smith, 1998).

$$\hat{y}_i = y_i - \varepsilon_i = b_0 + b_1 x_i \quad (2)$$

Levine *et al.* (2013) destacam que o b_0 e b_1 são determinados pelo Método dos Mínimos Quadrados (MMQ). A linha de regressão que possui o ajuste ótimo, segundo o critério do MMQ, é aquele em que a soma dos quadrados das diferenças entre os valores da variável dependente é minimizada, conforme Equação 3 (Draper; Smith, 1998). O valor de b_0 é dado pela Equação 4 (Draper; Smith, 1998).

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (3)$$

$$b_0 = \bar{y} - b_1 \bar{x} \quad (4)$$

sendo \bar{x} e \bar{y} , o valor médio de x e y , respectivamente.

A incerteza de b_1 , também chamada de erro padrão (*standard error* – SE), é dada pela Equação 5 (Draper; Smith, 1998).

$$SE = \sqrt{\frac{1}{n-2} \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (5)$$

Além da regressão linear simples supracitada, é possível utilizar bibliotecas de aprendizado de máquina com suporte a regressão linear. Nesta pesquisa serão utilizadas duas bibliotecas da linguagem Python:

- i. *Scikit-learn* (Pedregosa *et al.*, 2011) é uma das principais bibliotecas de propósito geral escritas em Python. Possui implementação de vários algoritmos de aprendizado de máquina, seja de regressão ou classificação, assim como de redes neurais. Além disso, seus algoritmos são acessíveis a não especialistas, sendo utilizada em múltiplas disciplinas e campos de aplicação (Abraham *et al.*, 2014). A biblioteca suporta regressão linear simples e múltipla, possuindo as classes “*LinearRegression*” e “*Ridge*” que permitem a regularização das regressões.
- ii. *Statsmodels* (Seabold; Perktold, 2010) é uma consolidada biblioteca estatística, amplamente empregada na área de econometria, análise de séries temporais, modelos lineares generalizados e regressão. Ademais, suporta uma ampla gama de classes, incluindo a OLS (*Ordinary Least Squares*), comumente utilizada para regressão linear.

Salienta-se que apesar das duas bibliotecas suportarem regressão linear, elas apresentam algumas diferenças entre as implementações e a forma de como lidar com certas configurações e otimizações, o que leva a variações nos resultados.

Fazilova, Ehgamberdiev e Kuzin (2018) destacam que o coeficiente angular do modelo de regressão linear que foi ajustado para cada série de coordenadas (X, Y e Z) corresponde às componentes das velocidade (v_x , v_y e v_z).

O MIDAS é um método robusto para a estimativa de velocidades a partir de séries temporais posicionais. Segundo Blewitt *et al.* (2016), trata-se de uma adaptação do estimador de Theil-Sen (Theil, 1950; Sen, 1968), capaz de fornecer estimativas de velocidade acuradas e insensíveis à sazonalidade. Entretanto, este método possui três características limitantes: fornece resultados de difícil interpretação caso a estação possua velocidades não uniformes; não mitiga efeitos sazonais que não sejam harmônicos anuais; e, não garante a robustez se a série temporal possuir dados ausentes. Mais informações sobre o MIDAS estão disponíveis em Blewitt *et al.* (2016).

O *Weighted Least Squares* (WLS) ou Mínimos Quadrados Ponderados, se baseia em três etapas: modelação dos dados - levando em consideração a presença de variância não

constante na série temporal; ponderação dos dados - tanto da variável de resposta quanto dos regressores; e, execução da regressão linear simples - combinada com métodos estatísticos para estimar os parâmetros do modelo e suas respectivas significâncias, tomando como base os dados ponderados. Romano e Wolf (2017) destacam que ponderar os dados resulta em ganhos de eficiência em relação ao estimador da regressão linear padrão .

O vetor de coeficientes estimados $\hat{\beta}_{WLS}$ é dado por Greene (2018) na Equação 6.

$$\hat{\beta}_{WLS} = (X^T W X)^{-1} X^T W y \quad (6)$$

sendo X^T a matriz transposta das variáveis independentes, W a matriz diagonal dos pesos w_i , e y o vetor das variáveis dependentes.

O erro padrão é dado pela Equação 7(Greene, 2018).

$$EP(\hat{\beta}_j) = \sqrt{\text{Elemento}(j, j) \text{ de } \hat{\sigma}^2 (X^T W X)^{-1}} \quad (7)$$

Minimum entropy (entropia mínima) é um método não paramétrico desenvolvido por Saleh, Bennett e Williams (2024) para estimar velocidades a partir de séries temporais posicionais, as quais podem conter ruídos, ausência de dados, *outliers* e descontinuidades.

O método toma como base a ideia de que a velocidade “ótima” de um objeto em movimento minimiza a entropia dos deslocamentos ao longo do tempo. Em outras palavras, entre todas as possíveis estimativas de velocidade, a que gera a distribuição mais "ordenada" (menos aleatória) dos resíduos é a mais provável (Saleh; Bennett; Williams, 2024).

O método parte de uma série temporal de posições do objeto em instantes de tempo discretos e suas respectivas coordenadas. Então defini-se um intervalo de busca - ajustando um modelo linear – e dividindo-o em incrementos de 0,01 mm/ano, no qual espera-se encontrar a velocidade “real”. Para cada valor de velocidade dentro do intervalo de busca, calcula-se a diferença entre as posições observadas e as posições teóricas previstas pelo modelo de velocidade constante. A entropia é então calculada para cada conjunto de resíduos. Por fim, a velocidade com menor entropia é considerada a velocidade “ótima” estimada (Saleh; Bennett; Williams, 2024).

Se a série contém descontinuidades, ela é dividida em sub-séries contínuas e a entropia total é calculada como uma média ponderada das entropias de cada sub-série, ignorando os saltos das descontinuidades.

Para estimar a incerteza da velocidade, a entropia mínima é convertida em um desvio padrão "efetivo" dos resíduos, assumindo que eles seguem uma distribuição Gaussiana. Esse desvio padrão é então usado para calcular a incerteza da velocidade, considerando o tempo total da série (Saleh; Bennett; Williams, 2024).

4.4 Identificação e remoção de *outliers*

Entre os métodos mais comuns utilizados para a suavização de séries temporais está a aplicação de médias móveis simples. Elas diminuem o ruído dos dados e destacam a tendência de longo prazo das séries (Shumway; Stoffer, 2011). Salienta-se que apesar de a média móvel não ser um método de detecção de *outliers*, ao suavizar as séries essas observações destoantes podem ser minimizadas.

Para calcular a média móvel simples, aplica-se uma janela móvel (*rolling window*) às séries e na sequência calcula-se a média aritmética dos dados contidos na janela. Matematicamente, a média móvel é dada pela Equação 8 (Morettin; Toloí, 2006).

$$MM_i = \frac{x_i + x_{i-1} + \dots + x_{i-n+1}}{n} \quad (8)$$

onde MM_i é a média móvel no índice i da série; e, n é o tamanho da janela.

Quando os dados a serem analisados - ou pelo menos certas características do conjunto desses dados - seguem uma distribuição normal, é possível utilizar técnicas de detecção de *outliers* baseadas no desvio padrão, como o *Z-score* e o teste dos três sigmas (3σ) (Ghilani, 2017). O desvio padrão mede a dispersão dos dados, basicamente indicando o quão distantes os pontos estão do valor médio. Para um conjunto de dados normal, espera-se que cerca de 68,2% dos dados estarão dentro de um desvio padrão da média, e 95,4% e 99,7% estarão respectivamente dentro de dois e três desvios padrão da média.

Dessa forma, para utilizar a técnica dos 3σ , define-se uma faixa entre três desvios padrão abaixo e três desvios padrão acima da média, da qual os pontos de dados que estejam fora são identificados como um *outlier*.

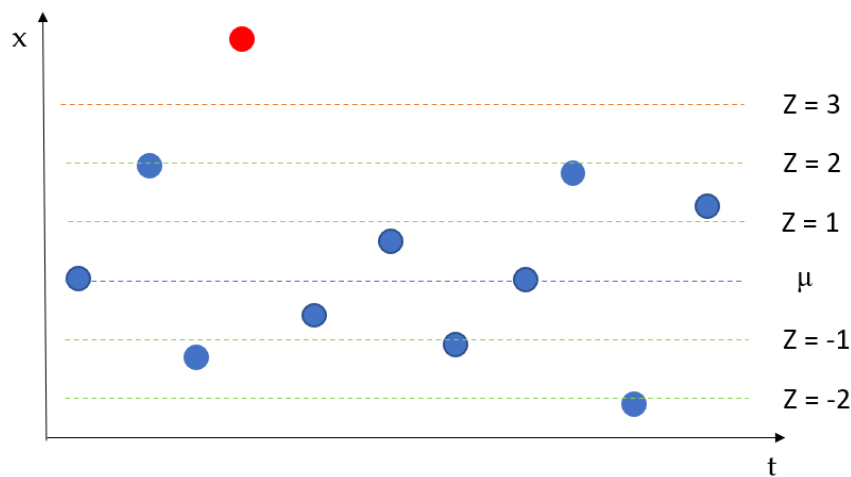
O *Z-Score* para um valor de x é calculado pela Equação 9 (Iglewicz; Hoaglin, 1993).

$$Z = \frac{(x-\mu)}{\sigma} \quad (9)$$

Observe que considerando uma faixa entre $(\mu - 3\sigma)$ e $(\mu + 3\sigma)$, temos $Z = -3$ e 3 . Logo esta técnica é equivalente à dos 3σ .

Para aplicação deste método define-se um *threshold* (limite), e calcula-se o *Z-Score* de todos os pontos da série, de forma que os valores maiores que o *threshold* são considerados *outliers*. A Figura 3 ilustra a aplicação do *Z-Score* com *threshold* igual a 3, onde o ponto detectado como *outlier* está em vermelho.

Figura 3 – Identificação de *outliers* com *Z-Score* em uma série temporal de coordenadas diárias (componente X).



Fonte: elaboração própria.

Morettin e Bussab (2004) destacam que o *Z-Score* possui dois pontos negativos relevantes: ineficiência quando empregado em dados não normais; e, dependência tanto da média como do desvio padrão, estatísticas sensíveis a valores extremos, e portanto, a *outliers*.

Alternativamente, tem-se o *Z-Score* modificado, em que a mediana (Q_2) substitui a média e o Desvio Absoluto da Mediana (MAD) substitui o desvio padrão (Iglewicz; Hoaglin, 1993). O *threshold* (M_i) para a identificação dos *outliers* é dado pela Equação 10:

$$M_i = \frac{0,6745 \cdot (x_i - Q_2)}{MAD} \quad (10)$$

A MAD é dada pela Equação 11:

$$MAD = \text{mediana} \{|x_i - Q_2|\} \quad (11)$$

sendo x_i o valor individual no conjunto de dados e $Q2$ a mediana dos dados.

A constante 0,6745 é utilizada de modo a normalizar a MAD, fazendo com que ela seja comparável ao desvio padrão em uma distribuição normal. Tal valor é aproximadamente igual ao inverso do quantil 75% de uma distribuição normal padrão. Desse modo, o *Z-Score* modificado pode ser interpretado de forma similar ao *Z-Score* tradicional, além de possuir a vantagem de ser mais robusto contra *outliers* e distribuições não normais.

Outro método estatístico comumente utilizado na detecção de *outliers* é o *boxplot* ajustado, baseado no *medcouple* (MC), que é uma medida de assimetria desenvolvida por Brys, Hubert e Struyf (2004). O MC é definido pela Equação 12:

$$MC(x_1, x_2, \dots, x_n) = \text{mediana} \left(\frac{(x_j - \text{med}_k) - (\text{med}_k - x_i)}{x_j - x_i} \right) \quad (12)$$

onde med_k é a mediana dos n valores de x ; e, i e j são subgrupos de x_n .

Os limites para a detecção dos *outliers* são dados pelas Equações 13 e 14 - adaptadas de Hubert e Vandervieren, (2008):

$$MC \geq 0 \rightarrow \begin{cases} LS = Q3 + 1,5 \cdot e^{\text{threshold} \cdot MC} \cdot AIQ \\ LI = Q1 - 1,5 \cdot e^{-(\text{threshold}+1) \cdot MC} \cdot AIQ \end{cases} \quad (13)$$

$$MC < 0 \rightarrow \begin{cases} LS = Q3 + 1,5 \cdot e^{(\text{threshold}+1) \cdot MC} \cdot AIQ \\ LI = Q1 - 1,5 \cdot e^{-\text{threshold} \cdot MC} \cdot AIQ \end{cases} \quad (14)$$

onde AIQ (Amplitude Interquartil Ajustada) é o intervalo interquartil, ou seja, $Q3 - Q1$ (terceiro quartil – primeiro quartil).

Salienta-se que Vandervieren e Hubert (2004) utilizam o *threshold* com valor de 3,5 e Hubert e Vandervieren (2008) com valor de 4.

Outro método estatístico, mas que foi recentemente proposto, é o “Método de Análise de Ponto de Mudança Bayesiana e Identificador de Hampel”, que alia dois métodos previamente desenvolvidos para detectar *outliers* em séries temporais posicionais.

A análise do Ponto de Mudança Bayesiano (BCP) é um método voltado para a localização de pontos de mudança (*change points* – CP) em um conjunto de dados (Grzechca

et al., 2017). Ele assume a existência de duas ou mais regiões com estruturas de distribuição distintas em uma série temporal e define matematicamente dois ou mais modelos de distribuição potenciais, sendo o ponto de mudança a fronteira entre duas regiões (Barry; Hartigan, 1993).

O Identificador de Hampel (*Hampel Identifier* - HI) determina prováveis *outliers* utilizando o valor da mediana para prever a localização do conjunto de dados e o MAD para estimar o desvio padrão (Dagar *et al.*, 2018; Pehlivan, 2024).

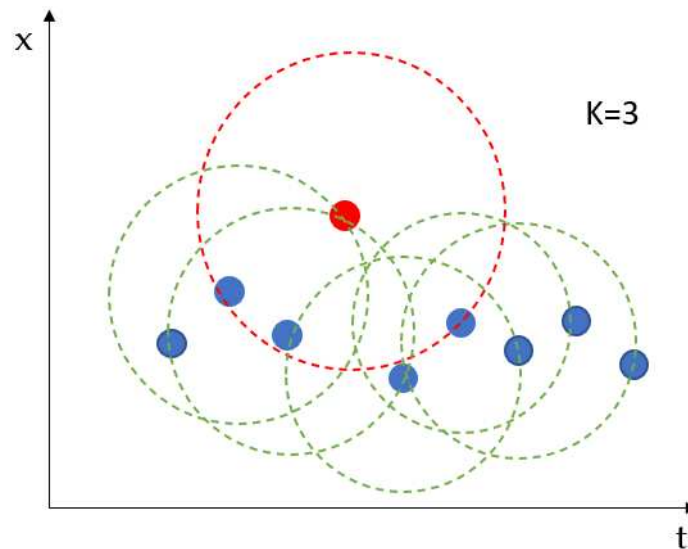
Com base nos dois métodos supracitados, Pehlivan (2024) desenvolveu o método *Bayesian Change Point analysis and Hampel Identifier* (BCP-HI). O BCP-HI identifica *outliers* com o método HI entre os potenciais pontos de mudança encontrados com o BCP. Desta forma, é possível identificar de forma eficaz anomalias verdadeiras e não falsos positivos, derivados de outros pontos de mudança ou de ruídos (Pehlivan, 2024). Mais detalhes sobre o método BCP-HI podem ser consultados em Pehlivan (2024).

Na sequência serão descritos três dos métodos mais comumente utilizados para a identificação de *outliers* baseados em aprendizado de máquina (Papastefanopoulos; Linardatos; Kotsiantis, 2021; Jain; Bajpai; Pamula, 2022).

O *K Nearest Neighbors* (KNN) (Ramaswamy; Rastogi; Shim, 2000) é um algoritmo supervisionado não-paramétrico frequentemente empregado para problemas de regressão e detecção de *outliers*. A suposição fundamental do KNN é que observações semelhantes estão próximas umas das outras, já os *outliers* geralmente são observações isoladas, afastadas do agrupamento de observações semelhantes.

O KNN mede as distâncias de um ponto para seus k vizinhos mais próximos e as utiliza como uma aproximação para quantificar a densidade dos dados naquela área. Após determinar os k vizinhos mais próximos de um ponto de dados, uma pontuação da probabilidade deste ponto ser *outlier* é produzida, geralmente combinando as distâncias calculadas com alguma função de agregação: a máxima, a média ou a mediana das k distâncias. A técnica de cálculo de distância mais empregada é da distância euclidiana (Papastefanopoulos; Linardatos; Kotsiantis, 2021). A Figura 4 ilustra o funcionamento do KNN para detectar *outliers* em uma série temporal posicional de dados diários (componente X). Considerando $k=3$, um ponto é classificado como *outlier* (ponto vermelho) por possuir as maiores distâncias para seus 3 vizinhos mais próximos, possuindo assim, a menor densidade local.

Figura 4 - Identificação de *outliers* com KNN em uma série temporal posicional de dados diários (componente X).



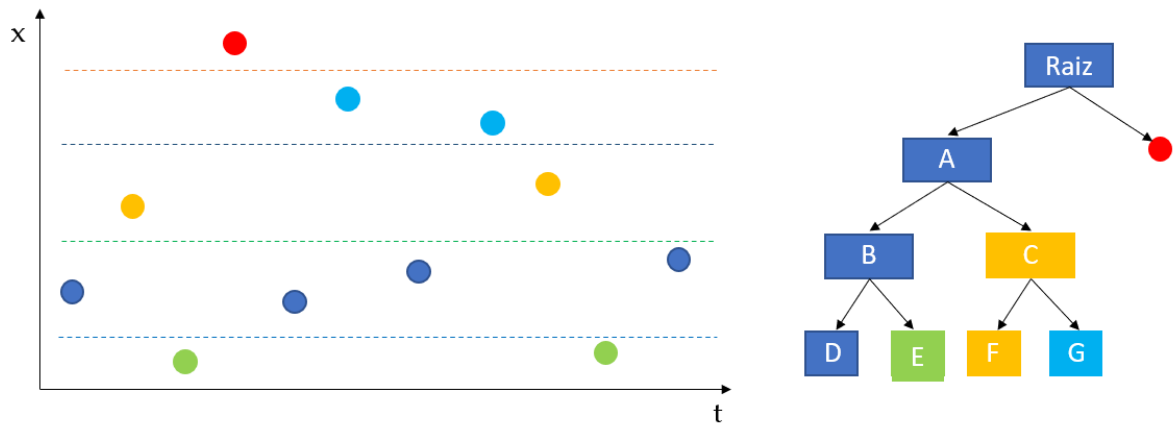
Fonte: elaboração própria.

Nota: adaptado de Lee, Park e Yoon (2020).

O *Isolation Forest* (IF), baseado no *Random Forest*, é um algoritmo não supervisionado que detecta pontos de dados anômalos. Para isso, ele considera duas suposições fundamentais: os *outliers* são uma minoria e incomuns. São criadas árvores de decisão com base em uma característica aleatória e, em seguida, é selecionada aleatoriamente uma regra de divisão para separar os valores com base na característica selecionada. Esse processo prossegue até que os valores dos parâmetros definidos sejam atingidos.

A principal métrica utilizada pelo algoritmo para analisar os dados é o comprimento do caminho necessário para isolar uma amostra, definido como a distância do nó raiz até o nó terminal correspondente. A média desse comprimento ao longo de uma floresta de árvores aleatórias serve como estimativa da “normalidade” da observação: comprimentos médios menores indicam maior probabilidade de anomalia. Dessa forma, a ocorrência recorrente de caminhos curtos para uma amostra, ao longo das diferentes árvores da floresta, indica com elevada probabilidade que tal instância seja uma anomalia no conjunto de dados (Liu; Ting; Zhou, 2008, 2012). A Figura 5 ilustra uma árvore aleatória de decisão para uma série temporal posicional de dados diários (componente X), onde o ponto com o menor caminho é provavelmente um *outlier* (ponto vermelho).

Figura 5 - Identificação de *outliers* com IF em uma série temporal posicional de dados diários (componente X).



Fonte: elaboração própria.

Nota: adaptado de Alam (2024).

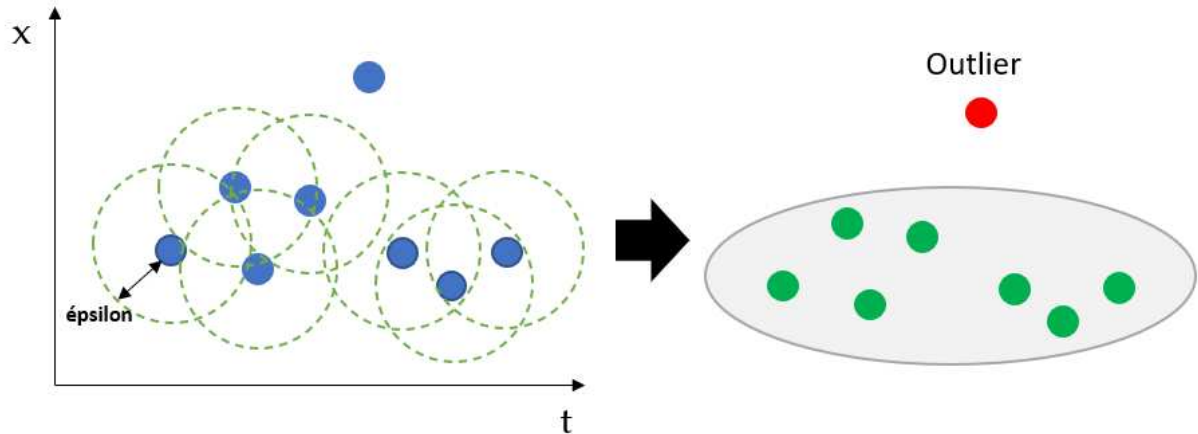
Uma das características mais vantajosas do IF é sua capacidade de tirar proveito de técnicas de amostragem, o que o torna um algoritmo muito rápido e com baixas exigências computacionais (Liu; Ting; Zhou, 2012).

O *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) é um método de agrupamento que se baseia na densidade das amostras. Ele agrupa os pontos que estão próximos uns dos outros - através da medição de distâncias Euclidianas, utilizando o parâmetro ϵ como *threshold* - e marca como *outliers* os pontos de baixa densidade que estão distantes (Ester *et al.*, 1996; Schubert *et al.*, 2017).

O DBSCAN funciona da seguinte forma: um ponto arbitrário não visitado previamente é selecionado e sua vizinhança é analisada. Caso o número de pontos nesta vizinhança seja maior que o número mínimo de pontos da amostra (o segundo parâmetro do método), um agrupamento é iniciado; do contrário, o ponto é marcado como *outlier*. Se posteriormente o ponto marcado como *outlier* fizer parte de uma vizinhança de tamanho adequado de um outro ponto, ele será incluído naquele agrupamento. Se um ponto estiver em uma zona de alta densidade de um agrupamento, então sua vizinhança também faz parte daquele agrupamento. Todos os pontos encontrados dentro da vizinhança são adicionados ao agrupamento, assim como sua própria vizinhança caso eles forem densos, até que se constate que o agrupamento conectado pela densidade está completo. Novamente, um ponto não visitado é analisado e processado conforme descrito acima, levando à determinação de um novo agrupamento ou

outlier (Jain; Bajpai; Pamula, 2022). A Figura 6 ilustra a identificação de *outliers* com DBSCAN em uma série temporal posicional de dados diários (componente X).

Figura 6 - Identificação de *outliers* com DBSCAN em uma série temporal posicional de dados diários (componente X).



Fonte: elaboração própria.

Nota: adaptado de Mantri (2021).

Segundo Belay *et al.* (2023), o DBSCAN é um algoritmo popular de agrupamento não supervisionado, sendo amplamente utilizado por poder lidar com amostras relativamente grandes.

Após a aplicação dos métodos de identificação e remoção de *outliers*, pode-se analisar algumas métricas estatísticas para verificar o ajuste e representatividade dos modelos de regressão linear aplicados nas séries, como as descritas a seguir.

O coeficiente de determinação (R^2), que avalia o ajuste de um modelo aos dados observados, indica a fração da variância da variável dependente que pode ser atribuída às variáveis independentes no modelo. O R^2 é dado pela Equação 15 (Wooldridge, 2016; Montgomery; Peck; Vining, 2021):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (15)$$

sendo n o número de observações, y_i os valores observados, \bar{y} a média das observações, e \hat{y}_i os valores preditos.

O R^2 ajustado, versão modificada do R^2 , é mais indicado para a comparação e avaliação de modelos de regressão devido ao ajuste pelo grau de liberdade, que penaliza a inclusão de variáveis independentes irrelevantes. O R^2 ajustado é dado pela Equação 16 (Wooldridge, 2016; Montgomery; Peck; Vining, 2021):

$$R_{ajustado}^2 = 1 - \left(\frac{(1-R^2)(n-1)}{n-p-1} \right) \quad (16)$$

sendo p a quantidade de variáveis independentes.

A interpretação do R^2 ajustado é dada por Wooldridge (2016) e Montgomery, Peck e Vining (2021):

- $R_{ajustado}^2 = 0$: modelo mal ajustado, as variáveis independentes não têm poder preditivo;
- $R_{ajustado}^2 = 1$: indica um ajuste perfeito;
- $0 < R_{ajustado}^2 < 1$: um valor mais próximo de 1 indica um melhor ajuste.

Outro modo de avaliar o ajuste do modelo é através do teste qui-quadrado, dado pela Equação 17 (Agresti, 2013; Casella; Berger, 2002).

$$X^2 = \sum \left(\frac{y_i - \hat{y}_i}{\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-p-1}}} \right)^2 \quad (17)$$

Se o valor de X^2 calculado é aproximadamente igual a $n - p - 1$, significa que modelo está bem ajustado e que os resíduos são independentes e distribuídos conforme esperado. Para validar o X^2 pode-se analisar o p -valor correspondente, que representa a probabilidade de observar um valor do teste igual ou superior ao valor calculado, dado que a hipótese nula seja verdadeira (Agresti, 2013; Casella; Berger, 2002). O p -valor é dado pela Equação 18:

$$p_valor = 1 - F_{x^2}(X_{calc}^2, GL) \quad (18)$$

onde $F_{x^2}(X_{calc}^2, GL)$ é a função de distribuição acumulada, X_{calc}^2 o valor calculado do qui-quadrado e GL é o número de graus de liberdade ($n - p - 1$).

A interpretação do p -valor é baseada em um nível de significância α , usualmente 5%:

- Se $p < \alpha$, rejeita-se a hipótese nula, indicando possível mau ajuste.
- Se $p \geq \alpha$, não há evidências estatísticas de mau ajuste.

Desta forma, utilizando o R^2 ajustado, o X^2 e o p -valor, pode-se verificar o ajuste dos modelos assim como verificar o impacto da remoção dos *outliers* das séries temporais posicionais.

4.5 Interpolação de velocidades

A estimativa de velocidades através de séries temporais posicionais possibilita a obtenção de velocidade de pontos conhecidos. Entretanto, para a obtenção de velocidades de pontos desconhecidos, pode-se realizar a interpolação dos pontos amostrais. Diversos métodos de interpolação podem ser utilizados, alguns dos mais utilizados serão descritos a seguir.

O *Griddata* é um método que interpola dados dispersos para um grid irregular, um grid regular, pontos dispersos, ou um conjunto de pontos específicos (NV5, 2024). Os pontos do grid são interpolados linearmente pela triangulação de Delaunay (Delaunay, 1934)

O *Inverse distance weighted* (IDW) estima valores desconhecidos calculando a média dos valores amostrados em suas respectivas vizinhanças, assumindo que regiões próximas são mais similares que as que aquelas estão mais distantes (Watson e Philip, 1985). Os valores amostrais que estão mais próximos dos pontos a serem estimados recebem um maior peso do que aqueles que estão mais distantes (ponderação inversa), e, portanto, exercem uma influência maior sobre o valor previsto (Munyati e Sinthumule, 2021).

O IDW é dado pela Equação 19 (Roberts *et al.*, 2004):

$$Z_{(x_0)} = \sum_{i=1}^n \lambda_i Z_{(x_i)} \quad (19)$$

onde $Z_{(x_0)}$ é o valor a ser interpolado, $Z_{(x_i)}$ é o valor amostral, n é o número de amostras e λ_i é o peso, dado por Roberts *et al.* (2004) (Equação 20):

$$\lambda_i = \frac{[d_{(x_i, x_0)}]^p}{\sum_{i=1}^n [d_{(x_i, x_0)}]^p} \quad (20)$$

sendo p o parâmetro de potência e $d_{(x_i, x_0)}$ a distância euclidiana entre x_i e x_0 e.

O parâmetro de potência controla a significância de pontos amostrais nos valores interpolados baseado na distância entre eles. Um valor de potência maior irá dar mais ênfase aos pontos amostrais mais próximos, fazendo com que eles tenham mais influência no valor interpolado (Watson e Philip, 1985). O IDW funciona bem com dados "ruidosos" e é mais adequado para amostras moderadamente densas, no que diz respeito à variação local (Akkala *et al.*, 2010).

A Interpolação por Funções de Bases Radiais (*Radial basis function interpolation - RBF*) é um método sofisticado na teoria da aproximação, utilizado para construir interpoladores de alta ordem a partir de dados não estruturados em espaços multidimensionais (Jakobsson; Andersson; Edelvik, 2009). Esse método fundamenta-se na combinação ponderada de funções de bases radiais, as quais são definidas como funções que dependem exclusivamente da distância entre o ponto de interesse e os pontos de dados. Estas funções são selecionadas para serem suficientemente "suaves" e "flexíveis" a fim de capturar a estrutura dos dados, sem serem excessivamente complexas (Du Toit, 2008).

O método RBF revela-se especialmente eficaz para abordar problemas de interpolação com dados escassos em espaços multidimensionais, onde a estrutura dos dados não é facilmente identificável por métodos tradicionais de interpolação que requerem uma malha predefinida. Ademais, este método é reconhecido por sua precisão e estabilidade, mesmo quando aplicado em conjuntos de dados volumosos (Du Toit, 2008; Chen *et al.* 2023). Formulação e detalhes sobre o RBF podem ser encontrados em Du Toit (2008), Jakobsson, Andersson e Edelvik (2009) e Chen *et al.* (2023).

O *Random Forest* (RF) é um algoritmo baseado na utilização de múltiplas árvores de decisão, geradas de forma aleatória a partir de diversos subconjuntos de treinamento. O RF utiliza uma técnica de amostragem chamada "agregação de *bootstrap*" (também chamada de "*bagging*"), na qual os dados para treinamento são amostrados de forma aleatória e com reposição, o que permite que as mesmas amostras sejam selecionadas várias vezes para a criação de diferentes árvores (Rodriguez-Galiano; Chica-Olmo; Chica-Rivas, 2014; Rodriguez-Galiano *et al.*, 2015).

À medida que as árvores são construídas, elas dividem iterativamente os nós raiz em nós folha binários e o algoritmo escolhe as melhores características dos dados para realizar as divisões, com o objetivo de separar os dados em grupos cada vez mais homogêneos em relação à variável de resposta. Essas divisões são realizadas até que uma condição de parada seja atingida, como uma profundidade máxima predefinida ou uma quantidade mínima de pontos em cada nó folha (Rodríguez-Galiano; Chica-Olmo; Chica-Rivas, 2014; Zhang; Zhou; Li, 2018). Em contraste com as árvores de decisão padrão, o RF seleciona aleatoriamente um subconjunto das variáveis preditoras de entrada para serem utilizadas como condições discriminativas em cada nó da árvore na floresta (SUN *et al.*, 2020).

Ao introduzir aleatoriedade tanto na seleção de amostras quanto na seleção de variáveis, o RF reduz a correlação entre as árvores individuais e aumenta a diversidade da floresta. Isso contribui para a robustez do algoritmo, ajudando a evitar o *overfitting* e a aumentar a qualidade das previsões (Breiman, 2001; Rodríguez-Galiano *et al.*, 2015).

Segundo Rodríguez-Galiano, Chica-Olmo e Chica-Rivas (2014), o RF é um método de aprendizagem de máquina com poderosas capacidades preditivas, com suas principais vantagens sendo:

- simplicidade da configuração de parâmetros;
- estimador imparcial e interno do erro de previsão;
- capacidade de trabalhar com dados complexos que possuem diferentes distribuições estatísticas, lidando com relações não lineares entre as variáveis.
- capacidade de utilizar preditores numéricos e categóricos; e,
- capacidade de determinar a importância das variáveis - o algoritmo atribui uma pontuação para cada variável, o que reflete sua contribuição nas previsões.

O *K-Nearest Neighbors* (KNN) é um método supervisionado que além de identificar *outliers*, como já descrito, também pode ser empregado em problemas de regressão (Song *et al.*, 2017). O algoritmo se baseia em instâncias (dados de entrada), o que significa que não há um modelo construído a partir de dados de treinamento. Em vez disso, todos os dados de treinamento são armazenados e, ao realizar uma regressão para uma nova instância, são feitos cálculos comparativos com os dados existentes. Ademais, por ser uma regressão não paramétrica, o KNN não faz nenhuma suposição sobre a distribuição dos dados (Goyal;

Chandra; Singh, 2014). O KNN determina a proximidade entre instâncias através do cálculo da distância entre os dados.

A krigagem ordinária (KO) é um método geoestatístico de interpolação espacial empregado para estimar valores não amostrados de uma variável espacialmente correlacionada (Journel; Huijbregts, 1978). Sua principal característica é assumir uma média local constante, mas desconhecida, na vizinhança de cada ponto que será estimado. Para cada ponto da vizinhança é atribuído uma ponderação derivada de um sistema de equações lineares que visa minimizar a variância do erro da estimativa e garantir que o estimador seja imparcial. Tal sistema de equações leva em consideração as distâncias entre os pontos amostrais e o ponto de interesse, assim como um modelo de variograma ajustado aos dados (Goovaerts, 1997; Yamamoto; Landim, 2013). Segundo Yamamoto e Landim (2013), entre os vários métodos de krigagem, o da krigagem ordinária é o mais amplamente empregado devido a sua simplicidade conceitual e a qualidade dos resultados gerados.

4.6 Atualização temporal de coordenadas

A atualização temporal das coordenadas de um ponto de interesse é dada pela Equação 21 (SIRGAS, 2025b):

$$\begin{aligned} X_{(t)} &= X_{(t_0)} + V_x(t - t_0) \\ Y_{(t)} &= Y_{(t_0)} + V_y(t - t_0) \\ Z_{(t)} &= Z_{(t_0)} + V_z(t - t_0) \end{aligned} \quad (21)$$

sendo t_0 a época de origem, t a época de destino; $X(t_0)$, $Y(t_0)$, $Z(t_0)$ são as coordenadas cartesianas na época t_0 ; $X(t)$, $Y(t)$, $Z(t)$ as coordenadas cartesianas na época t ; e V_x , V_y e V_z são as velocidades do ponto na época t_0 .

A propagação das incertezas no processo de atualização temporal das coordenadas é dada pelas Equações 22, 23 e 24 (Ramos; Dal Poz; Carvalho, 2022).

$$MVC_{\bar{X}(t)} = J_{AC} * MVC_{\bar{X}(t_0), \bar{V}} * J_{AC}^T \quad (22)$$

Sendo:

$$J_{AC} = \begin{bmatrix} 1 & 0 & 0 & (t - t_0) & 0 & 0 \\ 0 & 1 & 0 & 0 & (t - t_0) & 0 \\ 0 & 0 & 1 & 0 & 0 & (t - t_0) \end{bmatrix}_{3 \times 6} \quad (23)$$

$$MVC_{\vec{x}(t_0), \vec{v}} = \begin{bmatrix} MVC_{\vec{x}(t_0)} & 0 \\ 0 & MVC_{\vec{v}} \end{bmatrix}_{6 \times 6} \quad (24)$$

sendo $MVC_{\vec{x}(t)}$ a matriz variância e covariância das coordenadas atualizadas; J_{AC} a matriz jacobiana obtida após a derivada parcial da Equação 21 em relação aos seus parâmetros; e, $MVC_{\vec{x}(t_0), \vec{v}}$ a matriz variância e covariância das coordenadas na época t_0 . Salienta-se que todas as coordenadas e velocidades precisam estar associadas a um mesmo referencial.

4.7 Cálculo da discrepância planimétrica

De forma a calcular as discrepâncias planimétricas, inicialmente faz-se necessário transformar as coordenadas cartesianas (X, Y e Z) em geodésicas (ϕ , λ , h), associadas a um mesmo referencial. Petit e Luzum (2010) recomendam a utilização do elipsoide GRS80 na transformação.

O cálculo das discrepâncias planimétricas, em metros, é realizado conforme as Equações 25 a 30 (Torge, 2001).

Primeiramente, calcula-se, em radianos, as discrepâncias da latitude e longitude (Equação 25).

$$\begin{vmatrix} \Delta\phi(\text{graus}) \\ \Delta\lambda(\text{graus}) \end{vmatrix} = \begin{vmatrix} \phi(\text{referência}) - \phi(\text{estimada}) \\ \lambda(\text{referência}) - \lambda(\text{estimada}) \end{vmatrix} \quad (25)$$

Então calcula-se as discrepâncias em metros (Equação 26).

$$\begin{vmatrix} \Delta\phi(\text{metros}) \\ \Delta\lambda(\text{metros}) \end{vmatrix} = \begin{vmatrix} M \cdot \Delta\phi(\text{rad}) \\ N \cdot \cos\phi \cdot \Delta\lambda(\text{rad}) \end{vmatrix} \quad (26)$$

sendo M o raio de curvatura da seção meridiana (Equação 27) e N o raio de curvatura da seção primeiro vertical (Equação 28).

$$M = \frac{a(1-e^2)}{(1-e^2 \sin^2 \phi)^{\frac{3}{2}}} \quad (27)$$

$$N = \frac{a}{(1-e^2 \sin^2 \bar{\phi})^{\frac{1}{2}}} \quad (28)$$

onde a é o semieixo maior do elipsoide, $\bar{\phi}$ é a média entre as latitudes estimada e de referência, e e^2 é a segunda excentricidade (Equação 29).

$$e^2 = \frac{a^2 - b^2}{a^2} \quad (29)$$

sendo b o semieixo menor.

A discrepância planimétrica (T), é calculada através da Equação 30.

$$T_{(metros)} = \sqrt{\Delta_{\phi (metros)}^2 + \Delta_{\lambda (metros)}^2} \quad (30)$$

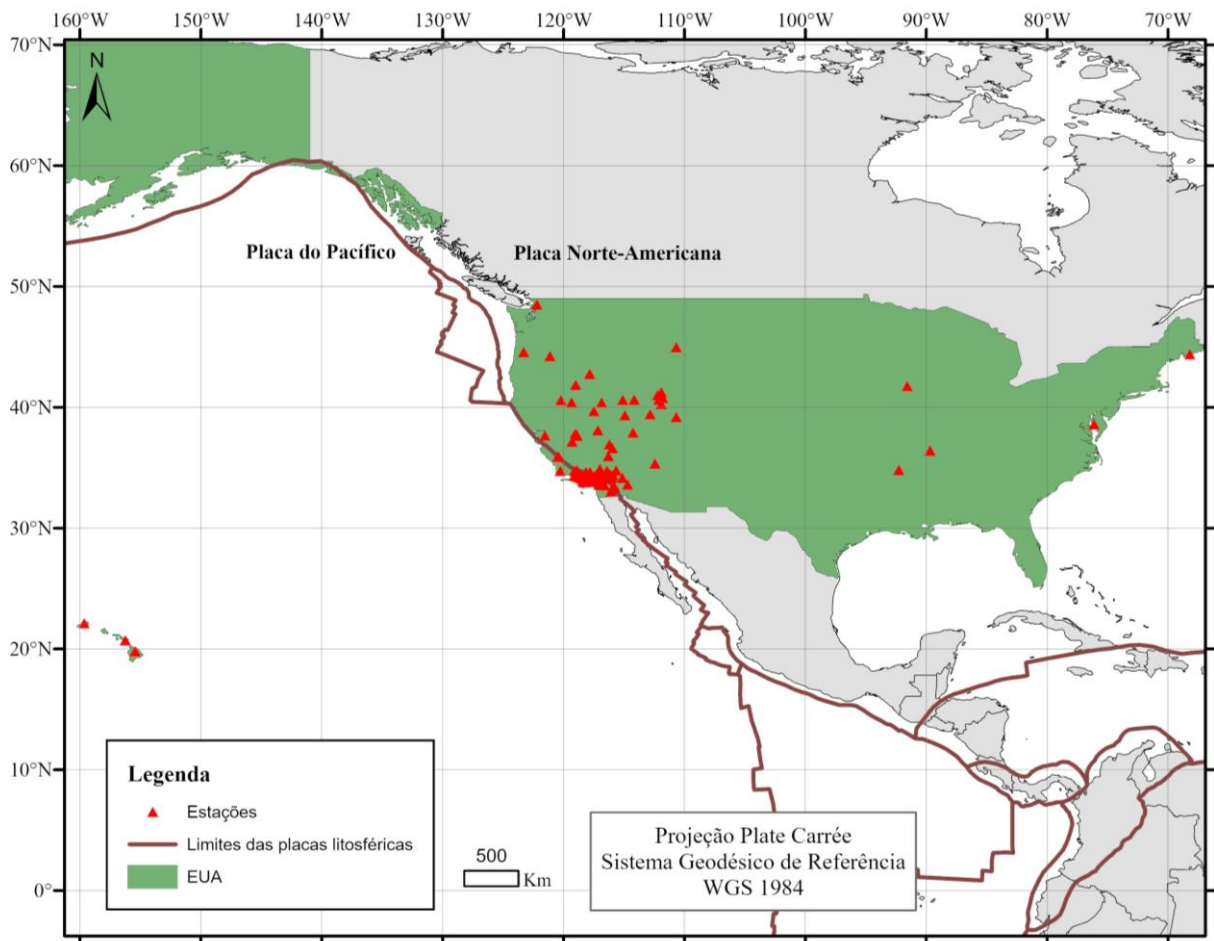
Como salientam Drewes e Heidbach (2012), a atualização temporal de coordenadas é um processo majoritariamente horizontal, portanto, as discrepâncias altimétricas normalmente não são analisadas.

5 MATERIAIS E MÉTODOS

5.1 Materiais

Para a realização dos experimentos foram utilizadas séries temporais posicionais (GPS-NGL) de 113 estações localizadas nos Estados Unidos da América (Figura 7), obtidas com o software desenvolvido: GPTS Downloader. A seleção das estações foi realizada de acordo com dois critérios: i) possuir série temporal que compreenda o período de 01/01/2000 a 31/12/2019 (20 anos); e, ii) possuir dados de observação RINEX disponíveis para seis dias no início da série – que serão utilizados para a estimativa das coordenadas de referência, como será discutido no item 5.2.6.

Figura 7 - Disposição das 113 estações no território dos Estados Unidos da América.



Fonte: elaboração própria.

Nota: disponível em alta qualidade em: bit.ly/Mapa-113https://bit.ly/Mapa_113.

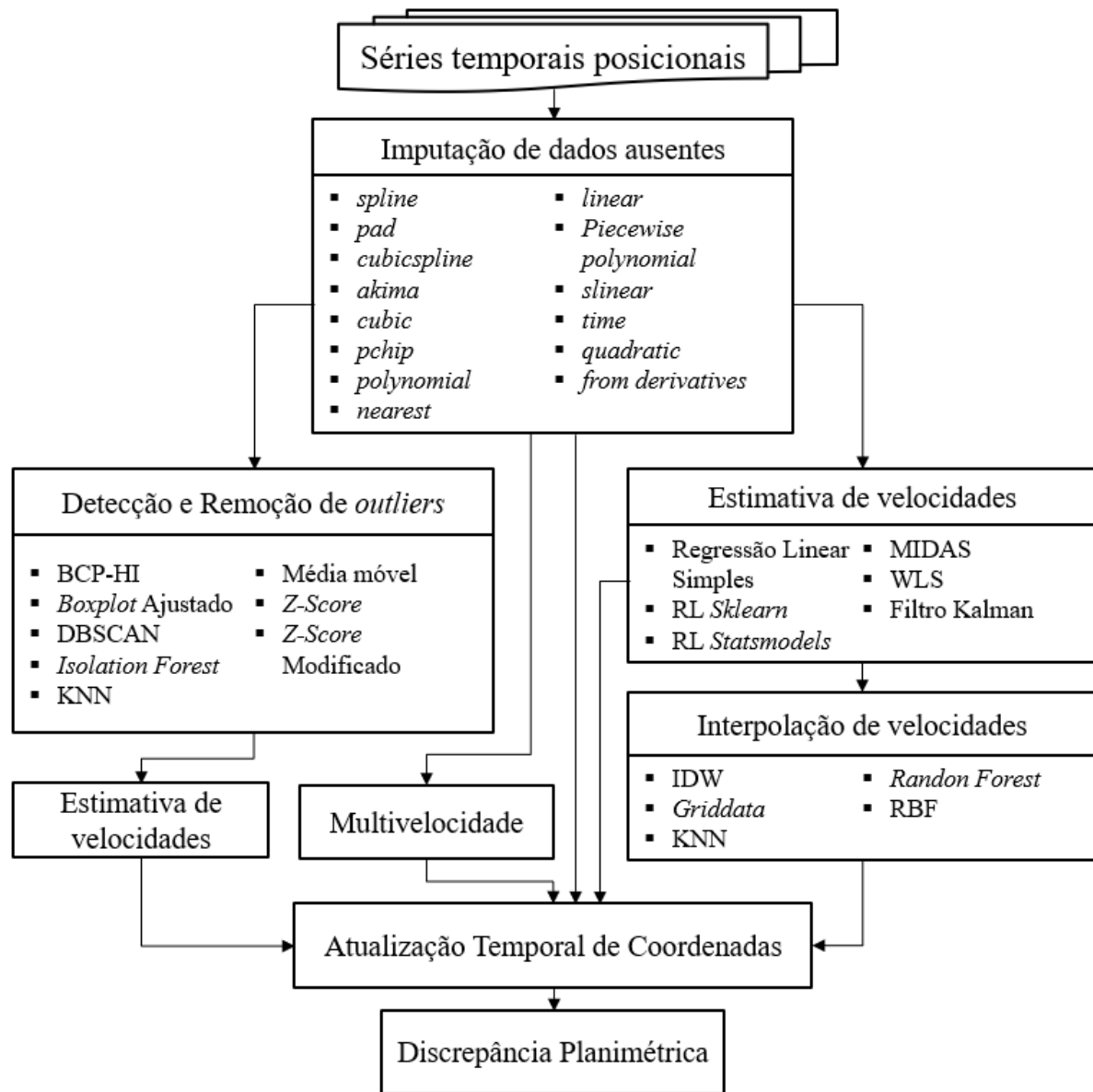
Uma tabela com os nomes, coordenadas, estado federativo e placa litosférica pertencente das 113 estações está disponível no APÊNDICE A.

5.2 Metodologia

Baseando-se no objetivo de otimizar a estimativa de velocidades para utilização no processo de atualização temporal de coordenadas, os métodos empregados serão analisados com base na discrepância planimétrica calculada após a realização dessa atualização.

O fluxograma geral do processo de atualização e as análises realizadas em cada etapa é exibido na Figura 8.

Figura 8 - Fluxograma da metodologia.



Fonte: elaboração própria.

Inicialmente, com a finalidade de automatizar o processo de obtenção das séries temporais posicionais, desenvolveu-se um software denominado GPTS Downloader.

5.2.1 GPTS Downloader

O software, desenvolvido em linguagem C, tem como objetivo facilitar tanto a seleção quanto a obtenção das séries temporais posicionais, possibilitando a classificação dos dados

quanto à data inicial e final de disponibilidade, período em anos decorrido entre essas duas datas, número de dias com dados, assim como a quantidade de dias com dados ausentes de todas as estações disponíveis. Adicionalmente, foram implementados filtros por localização que permitem a seleção de estações por países, além da possibilidade de o usuário inserir sua própria lista com estações específicas. A interface do GPTS Downloader pode ser visualizada na Figura 9.

Figura 9 –Interface do GPTS Downloader.

| Estação | Latitude (°) | Longitude (°) | h (m) | X (m) | Y (m) | Z (m) | Data Inicial | Data Final | Data Modificação | Nº dias com dados | Período (anos) | Dias sem dados |
|---------|--------------|---------------|---------|---------------|--------------|---------------|--------------|------------|------------------|-------------------|----------------|----------------|
| Q2NA | -12,4666 | -229,156 | 104,851 | -4073662,2759 | 4712064,7454 | -1367874,5086 | 27/03/2008 | 25/09/2018 | 25/09/2025 | 3185 | 10,50 | 643 |
| Q2NA | -12,4782 | -229,018 | 105,409 | -4094823,4607 | 4702026,6696 | -1369125,8893 | 08/04/2008 | 29/09/2019 | 25/09/2025 | 2360 | 11,48 | 1831 |
| Q2NA | -12,3559 | -229,1183 | 117,652 | -4078496,4388 | 4711300,1446 | -1355915,1775 | 22/09/2008 | 31/12/2016 | 25/09/2025 | 1913 | 08,28 | 1109 |
| QABI | 68,3543 | -341,1836 | 431,388 | 2233557,4175 | 761080,4273 | 5906106,1144 | 11/06/2009 | 27/09/2025 | 06/10/2025 | 5759 | 16,31 | 193 |
| QABN | 65,0337 | -338,6671 | 52,762 | 2514609,5568 | 982067,1548 | 5759343,7129 | 05/04/2023 | 27/09/2025 | 06/10/2025 | 879 | 02,48 | 27 |
| QABY | 58,6589 | -343,6204 | 60,548 | 3193913,9705 | 926687,3947 | 5424322,2456 | 10/04/2015 | 27/09/2025 | 06/10/2025 | 3790 | 10,47 | 33 |
| QAKE | 59,4814 | -341,6398 | 43,131 | 3082774,4229 | 1019646,6393 | 5471402,961 | 25/08/2010 | 27/09/2025 | 06/10/2025 | 5480 | 15,10 | 32 |
| QALE | 62,7805 | -343,9863 | 272,454 | 2811733,5415 | 806979,4457 | 5649069,3343 | 14/12/2015 | 27/09/2025 | 06/10/2025 | 3534 | 09,79 | 41 |
| QALF | 61,3445 | -343,3351 | 142,743 | 2946812,9614 | 848958,5594 | 5573978,8306 | 21/11/2005 | 27/09/2025 | 06/10/2025 | 7214 | 19,86 | 36 |
| QALI | 57,9295 | -347,4722 | 116,732 | 3313951,2113 | 736374,5033 | 5381673,7302 | 10/02/2011 | 27/09/2025 | 06/10/2025 | 5307 | 14,64 | 36 |
| QALM | 59,8663 | -341,9292 | 56,726 | 3051686,6597 | 995723,9513 | 5493063,2182 | 20/04/2002 | 27/09/2025 | 06/10/2025 | 8502 | 23,45 | 59 |
| QALN | 65,1321 | -338,9794 | 118,648 | 2510660,1077 | 964790,0082 | 5764026,5704 | 05/04/2023 | 27/09/2025 | 06/10/2025 | 874 | 02,48 | 32 |
| QALS | 56,9466 | -344,0928 | 158,208 | 3353530,0266 | 995732,8305 | 5322792,6315 | 19/10/2018 | 27/09/2025 | 06/10/2025 | 2509 | 06,95 | 26 |
| QALV | 61,231 | -345,9629 | 280,051 | 2985826,4894 | 746502,7274 | 59568019,7189 | 23/11/2005 | 19/04/2024 | 25/09/2025 | 6707 | 18,42 | 15 |
| QAMB | 60,5102 | -346,8488 | 192,096 | 3065281,6564 | 716200,4697 | 5528849,7358 | 09/05/2015 | 27/09/2025 | 06/10/2025 | 3766 | 10,39 | 28 |
| GAME | 58,8703 | -344,3998 | 143,433 | 3132908,0497 | 855548,4044 | 5436602,2471 | 16/07/2009 | 27/09/2025 | 06/10/2025 | 5855 | 16,21 | 62 |
| QAMM | 65,9582 | -343,7943 | 462,317 | 2502400,2137 | 727285,4144 | 5802472,3076 | 31/01/2008 | 27/09/2025 | 06/10/2025 | 6364 | 17,67 | 85 |
| QAMS | 64,5325 | -340,6501 | 248,871 | 2594858,6396 | 911244,4311 | 5735716,025 | 05/05/2016 | 27/09/2025 | 06/10/2025 | 3395 | 08,40 | 37 |

Fonte: elaboração própria.

Informações acerca das estações são exibidas em uma tabela personalizável na tela inicial do programa e é baseada no arquivo que contém a lista e metadados das estações disponíveis fornecido pelo NGL. Nesse sentido, o GPTS Downloader realiza automaticamente o *download* deste arquivo toda vez que é executado, fazendo com que as informações das estações estejam sempre atualizadas.

Em sua versão 1.0, o GPTS Downloader conta com filtros que separam as estações com base em sua localização. Por exemplo, ao selecionar o filtro do Brasil, no lugar de visualizar as mais de 22 mil estações espalhadas pelo mundo, serão mostradas apenas as 224 estações disponíveis no território brasileiro. Consultando as informações exibidas é possível verificar que dessas, 212 possuem pelo menos 365 dias de dados, enquanto 109 abrangem um período maior que 10 anos. A estação com maior período de atividade é a BRAZ, em operação há mais de 29 anos, com registros iniciados em 05/03/1995 e ainda em funcionamento. Entretanto, esta série apresenta 1138 dias sem dados, como mostrado na Figura 10.

Figura 10 - Recorte da interface do GPTS Downloader, utilização do filtro do Brasil e classificação das estações por maior período de dados.

| Estação | Data Inicial | Data Final | Data Modificação | Nº dias com dados | Período (anos) | Dias sem dados |
|---------|--------------|------------|------------------|-------------------|----------------|----------------|
| BRAZ | 05/03/1995 | 24/09/2024 | 07/10/2024 | 9658 | 29,58 | 1138 |
| UFPR | 14/12/1996 | 24/09/2024 | 08/10/2024 | 9676 | 27,80 | 470 |
| PPTE | 18/12/1996 | 28/09/2024 | 08/10/2024 | 9863 | 27,80 | 283 |
| POAL | 14/09/1999 | 24/09/2024 | 08/10/2024 | 8546 | 25,05 | 596 |
| CUIB | 21/01/2000 | 28/09/2024 | 07/10/2024 | 8643 | 24,70 | 374 |
| IMPZ | 21/01/2000 | 28/09/2024 | 07/10/2024 | 7229 | 24,70 | 1788 |
| CRA1 | 01/01/2001 | 21/09/2024 | 07/10/2024 | 6389 | 23,74 | 2275 |
| CHPI | 17/04/2001 | 17/09/2024 | 29/09/2024 | 7170 | 23,44 | 1384 |
| VICO | 25/04/2001 | 28/09/2024 | 08/10/2024 | 7080 | 23,44 | 1477 |
| SMAR | 01/01/2004 | 28/09/2024 | 08/10/2024 | 7256 | 20,76 | 320 |
| BELE | 01/01/2004 | 24/09/2024 | 07/10/2024 | 7085 | 20,75 | 487 |

Fonte: elaboração própria.

Salienta-se que os filtros são obtidos automaticamente utilizando a biblioteca *geopandas* (Jordahl *et al.*, 2020) da linguagem Python, através da qual é realizada uma seleção geoespacial das coordenadas das estações baseada em um *shapefile* de fronteiras políticas dos países - disponibilizado por Tapiquén (2015). Caso alguma estação de interesse não conste na lista do seu respectivo filtro, o usuário pode inserir sua própria lista de estações com a opção de “importar lista”, bastando para isso, colocar os nomes das estações em ordem alfabética em um arquivo de texto para ser importado. Esta opção também possibilita aos usuários selecionar estações específicas ou contidas em diferentes filtros para o *download* das séries posicionais, o que torna o processo de aquisição das séries mais eficiente.

Uma vez selecionadas as estações de interesse, basta clicar em um botão para realizar o *download* das séries posicionais. Há quatro formatos de arquivos disponíveis, que podem ser selecionados pelo usuário previamente ao *download*:

- i. o arquivo original das coordenadas cartesianas em notação científica, denominado {nome da estação}.txyz2.txt, contendo: nome da estação; data; ano decimal (época); X, Y, Z; X sigma (m); Y sigma (m); Z sigma (m); coeficiente de correlação XY, YZ e XZ; e altura da antena obtida do cabeçalho do arquivo RINEX;

- ii. o arquivo original de coordenadas (e , n , u) do SGL (Sistema Geodésico Local), denominado {nome da estação}.tenv.txt, contendo: nome da estação, data, ano decimal (época), dia juliano modificado, semana GPS, dia da semana GPS, e , n , u , altura da antena, e sigma, n sigma, u sigma, coeficientes de correlação en , eu , nu ;
- iii. arquivo no formato de entrada do software CoordHub, pertencente ao RINEXHub <rinexhub.wordpress.com>, denominado {nome da estação}.txyz2_CoordHub.txt, contendo: nome da estação, X, Y, Z, sistema de referência, data;
- iv. arquivo completo com todas as informações da série, denominado {nome da estação}.txyz2_completo.txt, sendo elas: nome da estação, X, Y, Z, sistema de referência, data da coordenada, X sigma (m), Y sigma (m), Z sigma (m), coeficiente de correlação XY, YZ e XZ, altura da antena, latitude, longitude e altitude geométrica; e,
- v. arquivo com todas as informações das velocidades MIDAS de todas as estações selecionadas, denominado Velocidades_MIDAS.txt.

Adicionalmente, há um aplicativo interno chamado “Interpolar Velocidades”, desenvolvido em Python, cuja funcionalidade consiste em ler todos os arquivos *completo.txt gerados pelo GPTS Downloader, calcular a velocidade para cada estação através de sua respectiva série temporal (por regressão linear simples) e interpolar as velocidades para os pontos de interesse inseridos pelo usuário. São disponibilizados resultados com os seguintes métodos de interpolação: *griddata*, *Inverse distance weighted* (IDW), *K-Nearest Neighbors* (KNN); *Radial basis function interpolation* (RBF) e, *Random Forest Regression* (RF).

O software GPTS Downloader está disponível de forma gratuita no endereço rinexhub.wordpress.com/gptsdownloader. Mais detalhes sobre o GPTS Downloader e sua operacionalização podem ser consultados no seu manual, disponível no menu “Ajuda”.

5.2.2 Imputação de dados ausentes

Todas as séries temporais posicionais utilizadas nesta pesquisa apresentaram ausência de dados para alguns dias ao longo do período de 20 anos. A imputação desses dados ausentes foi realizada utilizando 14 métodos de interpolação disponibilizados pela função *DataFrame.interpolate* da biblioteca Pandas da linguagem Python, sendo eles:

- | | | | |
|------|--------------------|-------|-----------------------------|
| i. | <i>spline</i> | viii. | <i>nearest</i> |
| ii. | <i>pad</i> | ix. | <i>linear</i> |
| iii. | <i>cubicspline</i> | x. | <i>Piecewise polynomial</i> |
| iv. | <i>akima</i> | xi. | <i>slinear</i> |
| v. | <i>cubic</i> | xii. | <i>time</i> |
| vi. | <i>pchip</i> | xiii. | <i>quadratic</i> |
| vii. | <i>polynomial</i> | xiv. | <i>from derivatives</i> |

De forma a selecionar o melhor método, foi realizado a imputação dos dados ausentes de todas as séries com todos os métodos e avaliado os valores das discrepâncias obtidas após atualização de coordenadas.

5.2.3 Identificação e remoção de outliers

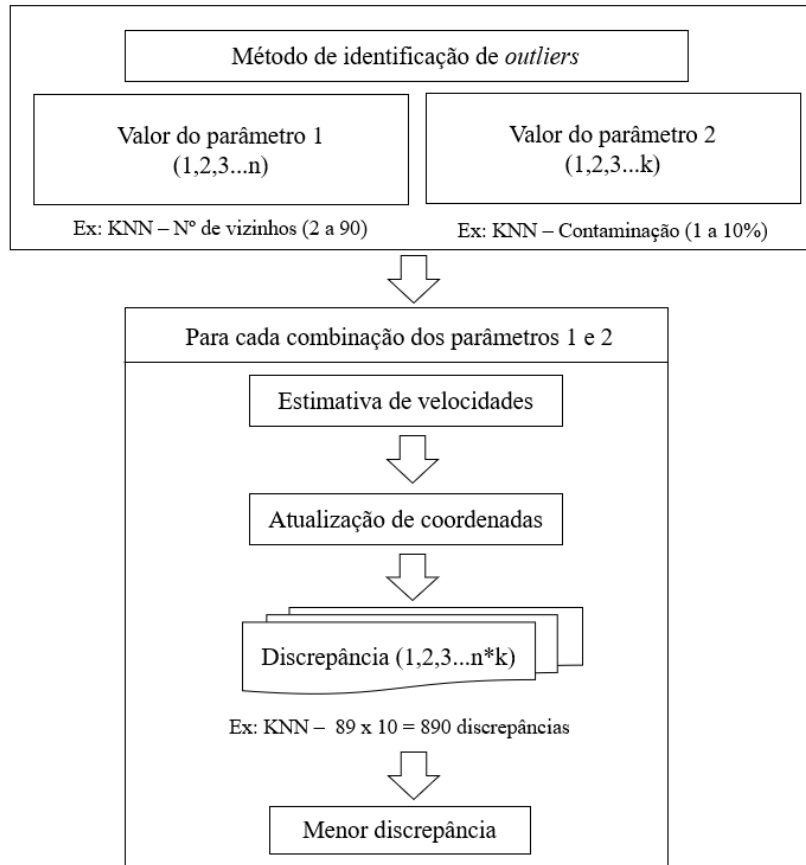
Excetuando-se a média móvel simples e o BCP-HI, os demais métodos analisados necessitam que os dados não possuam tendência para serem aplicados. Desta forma, para estes métodos, foi realizada a remoção da tendência das séries com a biblioteca *Statsmodels* da linguagem Python (Statsmodels, 2025). De forma semelhante foi realizada também a remoção da sazonalidade das séries, utilizando a função *seasonal_decompose*, também pertencente a biblioteca *Statsmodels*. Este último processo foi realizado visando comparar a performance dos métodos na presença e ausência da sazonalidade.

Destaca-se que tais procedimentos foram executados em cópias das séries, eliminando a necessidade de reinserir a tendência e a sazonalidade aos dados. Uma vez identificados os *outliers*, os índices destes pontos foram utilizados para remover as respectivas coordenadas na série original, que passou então pelo processo de imputação de dados ausentes novamente.

Os métodos de detecção de *outliers* utilizam parâmetros de configuração, como, por exemplo, número de vizinhos, *threshold* (limites) e contaminação (% de *outliers*). No entanto, não há um valor ótimo para estes parâmetros. Assim sendo, foi considerado um conjunto destes valores, visando analisar quais seriam mais indicados para aplicação em séries temporais posicionais. Para cada valor desse conjunto, foi realizado todo o processo de identificação e remoção de *outliers*, atualização das coordenadas e cálculo da discrepância. Por fim, foi selecionada apenas a menor discrepância, que foi utilizada para a discussão dos resultados.

Salienta-se que os métodos KNN e DBSCAN possuem dois parâmetros, o que resultou em um número significativamente maior de discrepâncias calculadas em relação aos demais métodos. Este processo é mostrado no fluxograma da Figura 11.

Figura 11 - Fluxograma da obtenção da menor discrepância com os métodos de identificação e remoção de *outliers*.



Fonte: elaboração própria.

Destaca-se que todos os métodos foram implementados em linguagem Python e que os intervalos selecionados para os parâmetros foram obtidos por embasamentos teóricos ou por resultados empíricos (caso se utilize um valor muito baixo para um parâmetro, nenhum *outlier* será detectado, e se o valor for muito alto, todos os pontos serão considerados *outliers* – dessa forma, é possível encontrar um intervalo com valores candidatos a serem utilizados na metodologia).

A média móvel simples foi empregada utilizando as funções `.rolling()` (Pandas, 2025b) e `.mean()` (Pandas, 2025c) da biblioteca Pandas. Foram avaliadas janelas de 2 a 90 dias (intervalo empírico), com incrementos unitários, totalizando 89 discrepâncias calculadas.

Para o *Boxplot* ajustado, foram utilizados os módulos “*scipy.stats*” e “*statsmodels.stats.stattools*”. Os valores de *threshold* testados compreendem o intervalo de 2 a 5, em incrementos de 0,5 (Vandervieren e Hubert (2004) utilizam 3,5 e Hubert e Vandervieren (2008) utilizam 4; empiricamente selecionou-se 2 a 5); totalizando 7 discrepâncias calculadas.

O *Z-Score* e o *Z-Score* modificado possuem apenas o parâmetro *Z-Score/threshold*, cujos valores analisados foram de 2 a 5, em incrementos de 0,5 (Blewitt *et al.* (2016) utilizam 2σ , Savchyn, Brusak e Tretyak (2022) utilizam o 3σ . Crocetti, Schartner e Soja (2021) utilizam 5σ); totalizando 7 discrepâncias calculadas.

O BCP-HI foi adaptado do código em MATLAB fornecido por Pehlivan (2024). Foram testados valores do parâmetro *cp* (*change points*) entre 2 e 100 (intervalo empírico associado ao utilizado por Pehlivan (2024)), em incrementos de 2, totalizando 50 discrepâncias calculadas.

Para o KNN foi utilizado a classe “*pyod.models.knn*” da biblioteca “PyOD” (Zhao; Nasrullah; Li, 2019). O método da classe selecionado foi o “*largest*”, que usa a distância para o *k*-ésimo vizinho como a pontuação de *outlier*. Foram calculadas 890 discrepâncias, uma vez que foram empregadas todas as combinações dos valores dos dois parâmetros: o valor da contaminação variando de 1 a 10% (Zhao; Nasrullah; Li, 2019), com incrementos de 1 ponto percentual; e o valor do número de vizinhos, variando de 2 a 90 (intervalo empírico), com incrementos unitários.

Para o IF foi utilizado a classe “*sklearn.ensemble.IsolationForest*” da biblioteca *scikit-learn* (Scikit-Learn, 2025a), e foram utilizados valores de contaminação de 1 a 10% (Liu; Ting; Zhou, 2008), com incrementos de 1 ponto percentual.

Para o DBSCAN empregou-se a classe “*sklearn.cluster.DBSCAN*” da biblioteca *scikit-learn* (Scikit-Learn, 2025b). Assim como o KNN, este método possui dois parâmetros: *épsilon* – cujos valores analisados foram de 0,0001 a 0,0015 (intervalo empírico), com incrementos de 0,0002; e o número mínimo de pontos da amostra – cujos valores empregados foram entre 2 e 90 (intervalo empírico), com incrementos unitários. O número total de discrepâncias calculadas foi 712.

5.2.4 Estimativa de velocidades

Para estimar a velocidade a partir das séries temporais posicionais, foram selecionados seis métodos: MIDAS; regressão linear simples; regressão linear simples *Scikit-learn*; regressão

linear simples *Statsmodels*; *minimum entropy* (ME); e, mínimos quadrados ponderados (*Weighted least squares* - WLS).

O MIDAS e o ME estão implementados e disponíveis no software SARI (Santamaría-Gómez, 2019), acessível no endereço <https://alvarosg.shinyapps.io/sari>. A metodologia empregada pelo SARI pode ser consultada em Santamaría-Gómez (2019).

Salienta-se que as velocidades foram estimadas diretamente das séries do NGL sem pré-processamentos para remoção de *outlier*, sazonalidade ou ruídos. Ramos, Dal Poz e Carvalho (2021) analisando o impacto da sazonalidade na estimativa de velocidades de estações brasileiras verificaram que negligenciar este aspecto das séries não causa alterações significativas nos resultados planimétricos. Adicionalmente, destacam a presença de ruídos brancos e ruídos de cintilação (ruído aleatório cuja potência é proporcional ao inverso da frequência) nas séries das estações não afetaram as velocidades estimadas (para séries com período de sete anos). Entretanto, Klos, Bos e Bogusz (2018) observam que se sinais sazonais e periodicidades residuais não forem propriamente removidos, as incertezas das velocidades serão artificialmente superestimadas.

Para avaliar o desempenho das velocidades estimadas através de séries temporais posicionais, foi realizada a comparação das discrepâncias planimétricas calculadas utilizando as velocidades estimadas e as discrepâncias calculadas com as velocidades fornecidas por 18 modelos de velocidades utilizados pela comunidade geodésica, disponibilizados pelo *Plate Motion Calculator* (UNAVCO, 2025):

- NUVEL 1 (Argus; Gordon, 1991);
- NUVEL 1A (Demets *et al.*, 1994);
- HS2-NUVEL1A (*Hot Spot2 Northwestern University Velocity model 1A*) (Gripp; Gordon, 1990; Demets *et al.* 1994);
- APKIM2000.0 (Drewes, 1998; Drewes; Angermann, 2001);
- ITRF2000 (Drewes; Angermann 2001);
- ITRF2000 (Altamimi; Sillard; Boucher, 2002);
- HS3-NUVEL1A (*Hot Spot3 NUVEL 1A*) (Gripp; Gordon, 2002);
- GSRM v1.2 (*Global Strain Rate Model*) (Kreemer; Holt; Haines, 2003);
- CGPS 2004 (*Continuous GPS 2004*) (Prawirodirdjo; Bock, 2004);
- REVEL 2000 (Sella; Dixon; Mao, 2004);
- APKIM2005-DGFI (*Deutsche Geodätische Forschungsinstitu*) (Drewes, 2009);

- APKIM2005-IGN (*Institute Géographique National*) (Drewes, 2009);
- MORVEL 2010 (*Mid Ocean Ridge Velocity 2010*) (Demets; Gordon; Argus, 2011);
- GEODVEL 2010 (*Geodesy Velocity 2010*) (Argus *et al.* 2010);
- NNR-MORVEL56 (Argus; Gordon; Demets, 2011);
- ITRF2008 (Altamimi; Collilieux; Métivier, 2012)
- GSRM v2.1 (Kreemer; Blewitt; Klein, 2014); e,
- ITRF2014 (Altamimi *et al.*, 2016).

Para o processo de atualização temporal, foram selecionadas as coordenadas referentes a todo o ano de 2019 (01/01 a 31/12), as quais foram atualizadas para o dia inicial da série (01/01/2000, época 2000,00). De forma a minimizar efeitos sazonais - como por exemplo, períodos de cheia e secas que causam alterações nas posições de estações localizadas próximas de grandes corpos d'água (Michel *et al.*, 2021; Laroche *et al.*, 2022) - optou-se por analisar a média das 365 discrepâncias obtidas para cada estação.

Destaca-se que como as séries temporais posicionais obtidas com o GPTS Downloader possuem todas as coordenadas em um mesmo referencial (IGS14/ITRF2014 – salienta-se que os dados foram obtidos antes de agosto de 2025, quando as séries passaram a ser referenciadas ao IGS20/ITRF2020), não se faz necessário a compatibilização de referencial nem das coordenadas nem das velocidades - já que estas possuem o mesmo referencial das coordenadas das quais foram estimadas.

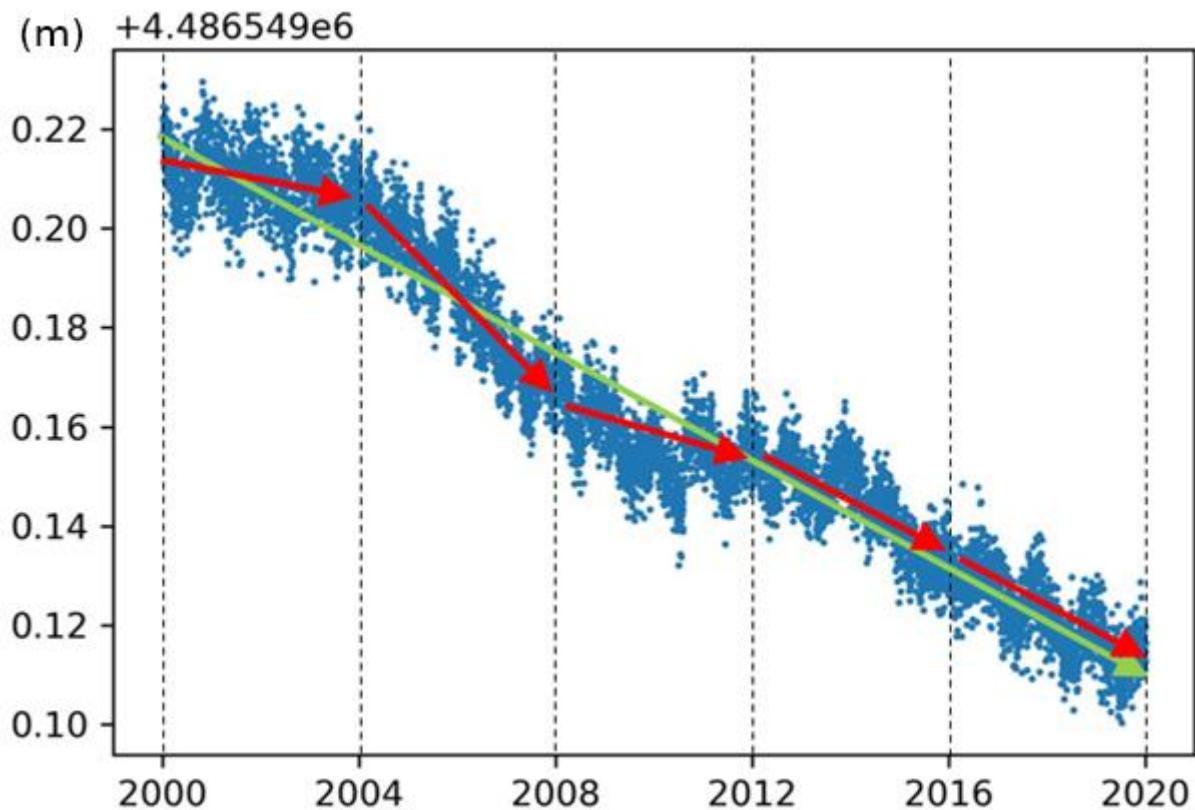
5.2.5 Multivelocidade

No processo de atualização temporal de coordenadas adota-se monovelocidades para os pontos de interesse, isto é, velocidades únicas e constantes no tempo. Entretanto, assim como inicialmente proposto em Freitas, Dal Poz e Nascimento (2022), avaliou-se a realização deste processo utilizando multivelocidades, ou seja, para um mesmo ponto atribui-se múltiplas velocidades, com cada uma associada a um trecho distinto do período de atualização.

Foram analisadas multivelocidade de 1, 2, 3, 4, 5, 6, e 10 anos. A seguir será abordado o procedimento para a estimativa de multivelocidade de 4 anos, que é análogo para os demais períodos, variando apenas o período de cada parte a qual a série será segmentada e o número de partes.

Para estimar as multivelocidades de 4 anos de um período de 20 anos, inicialmente divide-se a série em segmentos com períodos de 4 anos a partir da época inicial e então estima-se a velocidade de cada segmento. Destaca-se que a monovelocidade de 20 anos também foi calculada para comparação de desempenho na atualização das coordenadas. As multivelocidades e a monovelocidade são ilustradas na Figura 12.

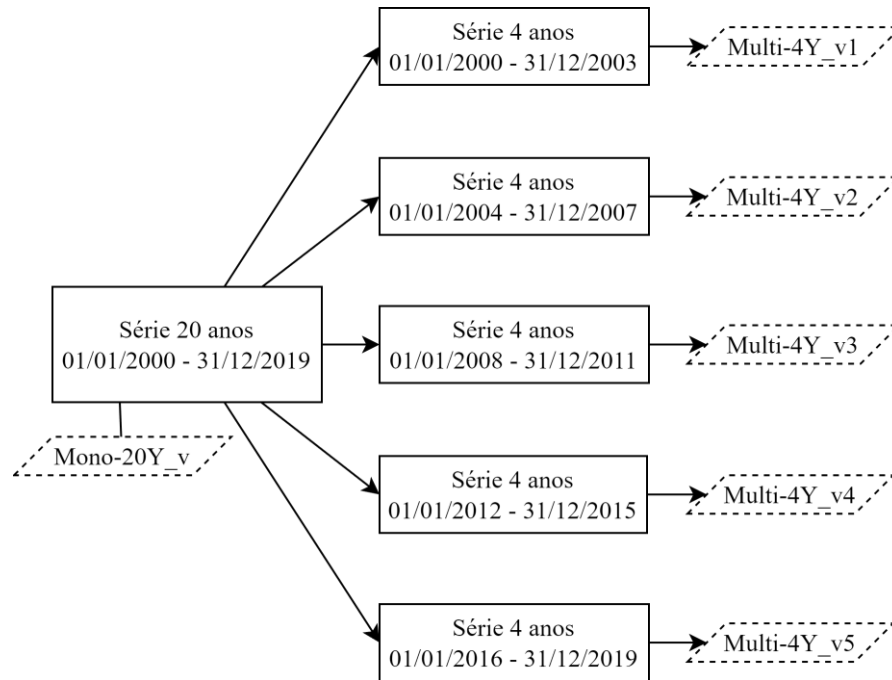
Figura 12 - Multivelocidades de 4 anos (setas vermelhas) e monovelocidade de 20 anos (seta verde) para a série de coordenada Z da estação MAWY.



Fonte: elaboração própria.

O processo de estimativa das velocidades de cada segmento de 4 anos está esquematizado na Figura 13.

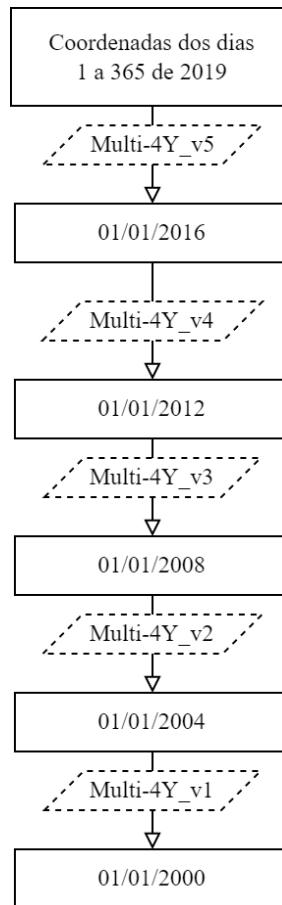
Figura 13 - Fluxograma do processo de estimativa das multivelocidades.



Fonte: elaboração própria.

Uma vez estimadas todas as velocidades, prossegue-se então para o processo de atualização temporal. As coordenadas utilizadas para atualização foram as mesmas empregadas na monovelocidade descritas no item anterior, isto é, as 365 coordenadas do ano de 2019. Tais coordenadas foram inicialmente atualizadas para 01/01/2016, empregando a multivelocidade 4Y_v5; na sequência foram atualizadas para 01/01/2012, empregando a 4Y_v4; então para 01/01/2008, empregando a 4Y_v3; então para 01/01/2004, empregando a 4Y_v2, e finalmente foram atualizadas para 01/01/2000, empregando a 4Y_v1. Este processo é ilustrado pelo fluxograma da Figura 14.

Figura 14 - Atualização das coordenadas referentes ao ano de 2019 para 2000,0, utilizando multivelocidades.



Fonte: elaboração própria.

De posse das coordenadas atualizadas para *dr/er* realiza-se então o cálculo da discrepância planimétrica em relação à coordenada de referência de cada estação. Essas discrepâncias foram então comparadas com aquelas obtidas com as monovelocidades.

Adicionalmente, foi analisado também a utilização de multivelocidades não uniformes, ou seja, segmentou-se uma mesma série em intervalos de duração distintos, de forma a conseguir uma melhor adequação ao deslocamento da estação.

5.2.6 Coordenadas de referência

Uma vez que não foram encontradas coordenadas oficiais de referência para a data de interesse - dia 01/01/2000, data de início das séries -, optou-se por processar arquivos de observação de 144 horas pelo método PPP no Canadian Spatial Reference System (CSRS-PPP), como recomendado por Ebner e Featherstone (2008) e adotado por Cunha (2020) e Freitas, Dal

Poz e Nascimento (2022). Para cada estação, concatenaram-se seis arquivos diários obtidos no GAGE Data File Server (GAGE, 2025) - três dias antes e dois dias após o dia 01/01/2000 -, garantindo que a data de referência ficasse próxima ao meio do intervalo processado.

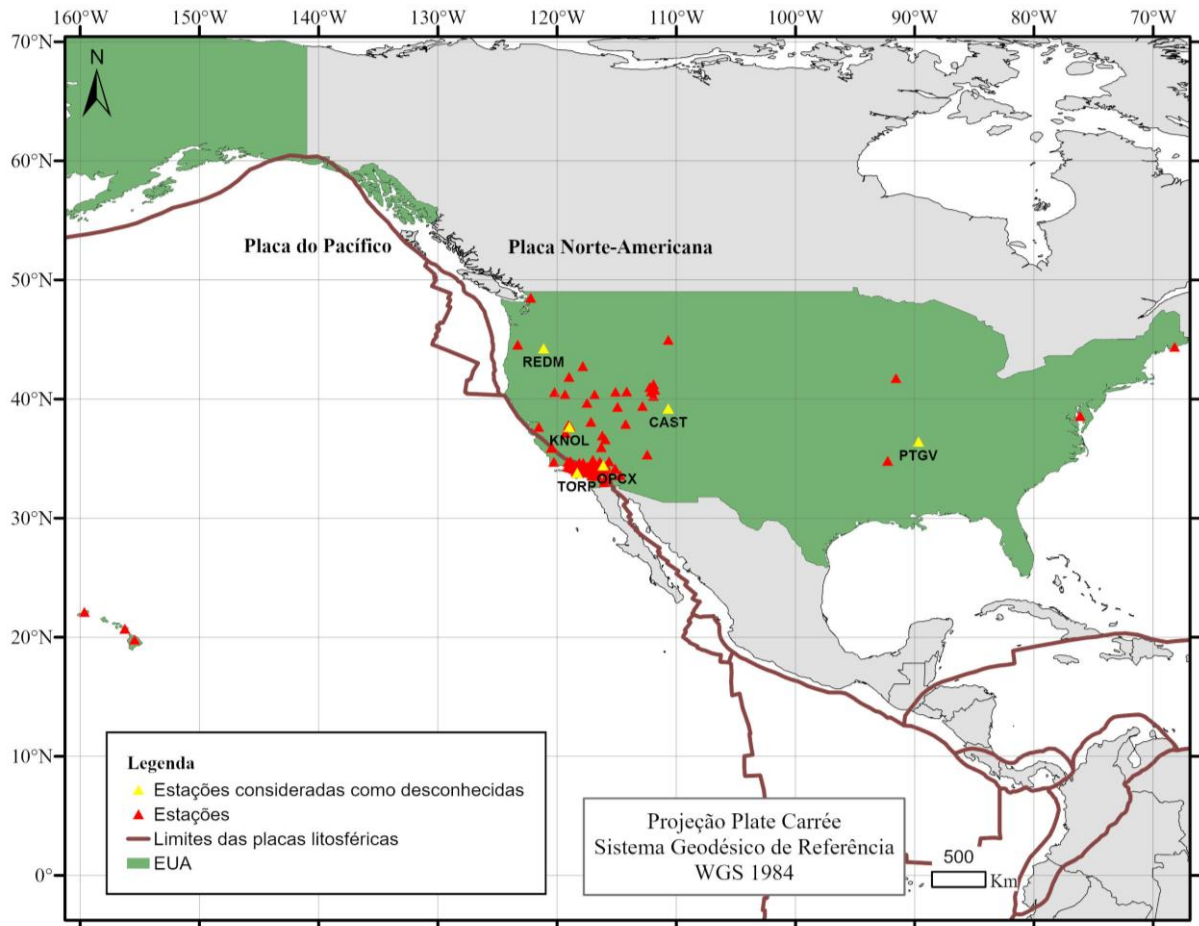
5.2.7 Interpolação de velocidades

Uma vez estimada as velocidades das estações, foi avaliada a utilização de cinco métodos de interpolação para obtenção de velocidades de pontos desconhecidos, disponíveis pelo aplicativo “Interpolar velocidades” do GPTS Downloader. Sendo eles:

- *Griddata* - função *griddata()* da biblioteca *SciPy*;
- IDW – algoritmo de implementação própria, cuja ponderação utilizada foi o inverso do quadrado da distância entre os pontos conhecidos e os pontos a serem interpolados. O parâmetro de potência utilizado foi igual a 2;
- KNN – função *KNeighborsRegressor()* da biblioteca *Scikit-learn*;
- Krigagem Ordinária (KO) – função *OrdinaryKriging* da biblioteca *pykrige*;
- RBF - função *rbg()* da biblioteca *SciPy*; e,
- RF - função *rf.predict()* da biblioteca *Scikit-learn*;

Com a finalidade de avaliar os resultados, seis das 113 estações foram consideradas desconhecidas (CAST, KNOL, PTGV, REDM, OPCX e TORP) e, portanto, excluídas do conjunto de entrada. As 107 estações restantes foram utilizadas como dados de entrada nas interpolações, e os valores de velocidade estimados para cada uma das seis estações foram comparados às velocidades calculadas a partir das respectivas séries temporais

Figura 15 - Localização das estações selecionadas para serem consideradas como desconhecidas.



Fonte: elaboração própria.

Nota: disponível em alta qualidade em: bit.ly/Mapa-6.

6 RESULTADOS

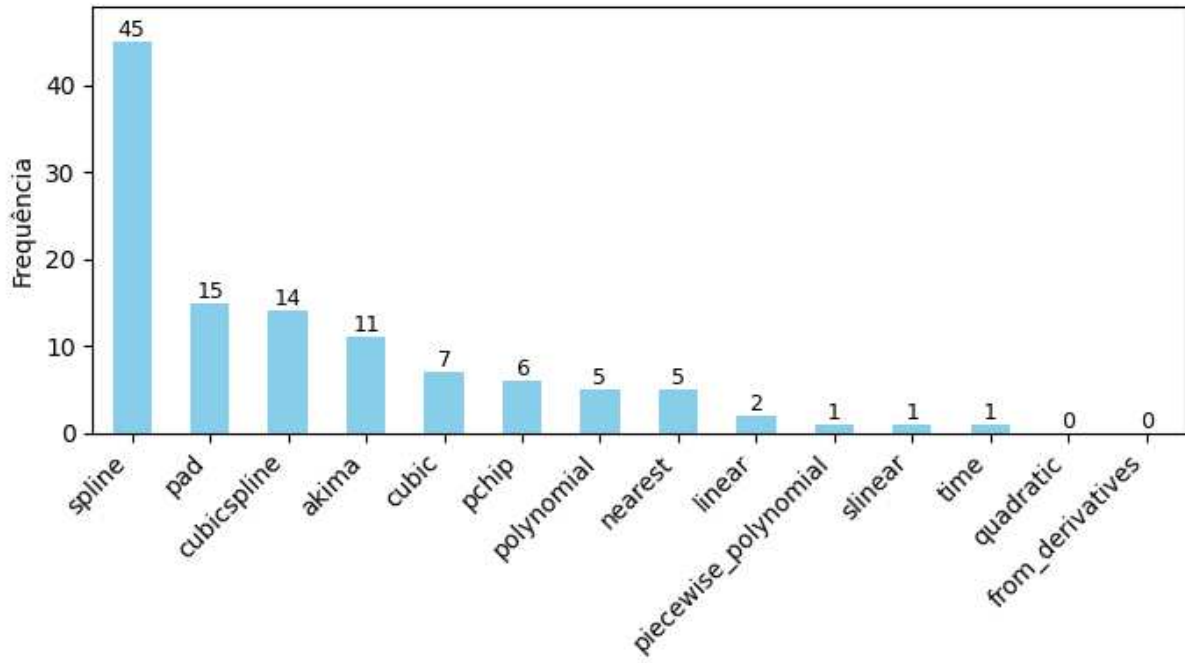
Devido ao grande número de séries e experimentos realizados, decidiu-se não incluir todos os gráficos ou tabelas completas com informações de todas as estações no corpo do texto. No entanto, todos os resultados gerados (mais de 5.600 gráficos), referentes a todas as estações e experimentos, estão disponíveis no endereço bit.ly/Resultados_Krisley.

6.1 Imputação de dados ausentes

Após imputar os dados ausentes utilizando cada um dos 14 métodos de imputação em cada uma das 113 séries, analisou-se qual método apresentou a menor discrepância para cada

estação. Como mostrado na Figura 16, a imputação utilizando *spline* apresentou menor discrepância em 45 das 113 estações, desempenho amplamente superior aos demais.

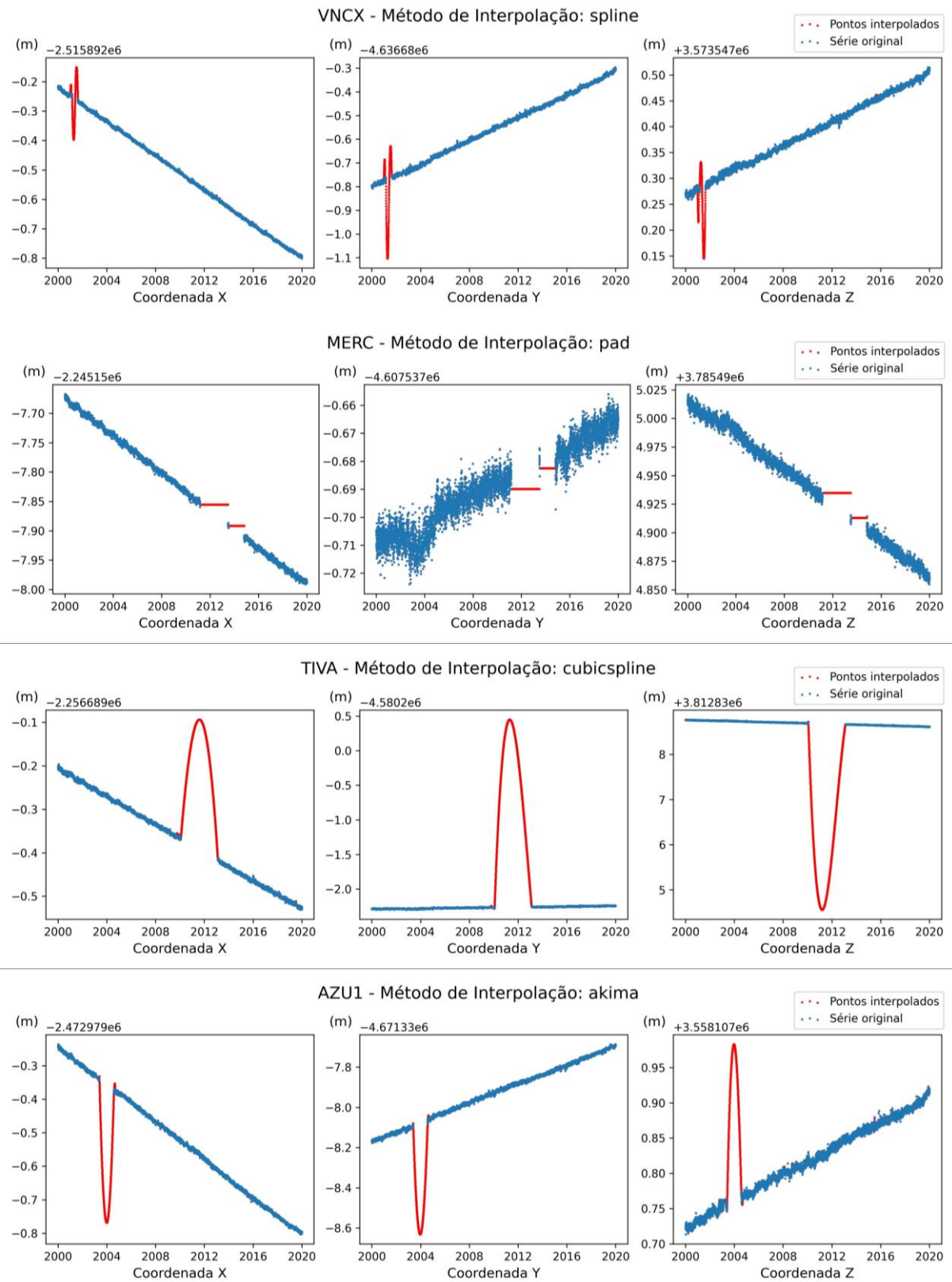
Figura 16 - Frequência de melhor discrepância por método para as 113 estações.



Fonte: elaboração própria.

Na sequência, destaca-se *pad*, *cubicspline* e *akima*, com os melhores resultados em 15, 14 e 11 estações, respectivamente. Entretanto, ao verificar graficamente a imputação dos dados ausentes, constatou-se que nas séries em que *spline*, *cubicspline* e *akima* não foram os melhores, apresentaram resultados significativamente piores. E mesmo nas séries em que tiveram bons resultados, a imputação não foi representativa do comportamento dos dados. Já o *pad* (assim como o *nearest*), apesar de proporcionar valores de discrepância menores em relação a alguns métodos, sempre realiza a imputação com o mesmo valor para todos os dias ausentes de uma mesma falha, não levando em consideração a tendência da série – o que resulta em uma linha horizontal. Esses problemas são ilustrados na Figura 17.

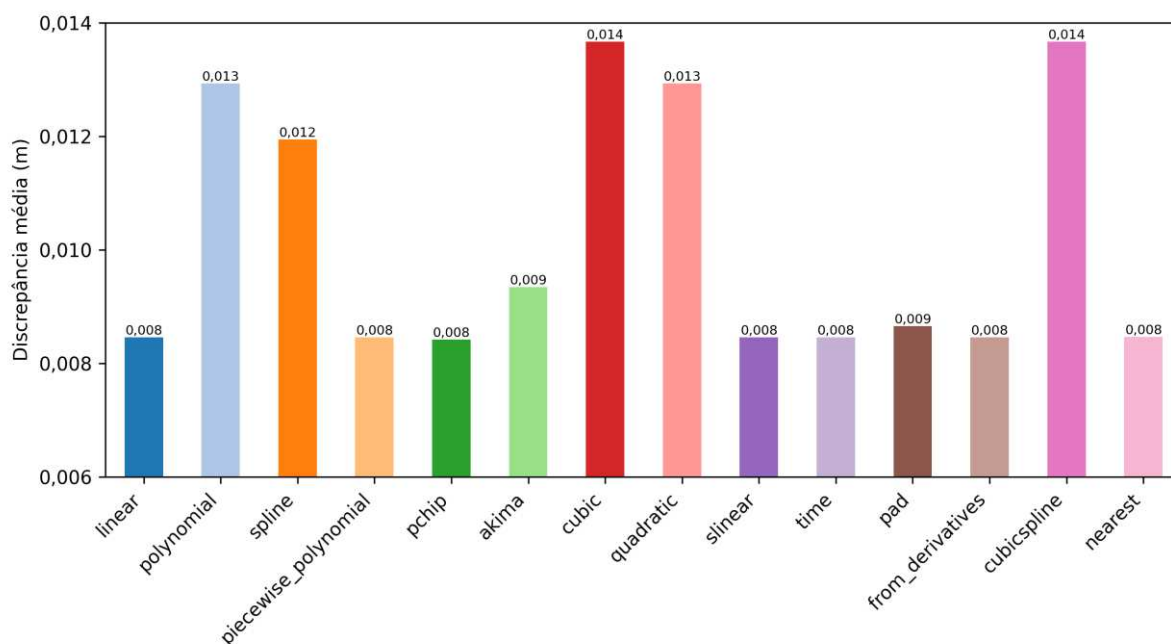
Figura 17 - Séries temporais posicionais com dados preenchidos com os métodos *spline*, *pad*, *cubicspline* e *akima*.



Fonte: elaboração própria.

Desse modo, estes métodos precisam de um emprego mais cuidadoso, necessitando que cada aplicação seja analisada particularmente. Como esta pesquisa se baseia na utilização de muitas séries, decidiu-se analisar as estações de forma conjunta, comparando a média das 113 estações com cada método, como exibido na Figura 18.

Figura 18 - Discrepância média das 113 estações para cada método.



Fonte: elaboração própria.

Os métodos *piecewise polynomial*, *linear*, *pchip*, *slinear*, *time*, *from derivatives* e *nearest*, apresentaram a menor discrepância média, de 0,008 m. No entanto, o *polynomial* ainda apresentou algumas imputações irregulares, o *nearest* apresenta o mesmo comportamento do *pad* e o *from derivatives* não apresentou melhor resultado em nenhuma estação.

Como os demais métodos apresentaram valores e comportamento semelhantes, optou-se pela utilização do método *linear*, já que foi consistente em todas as séries e é comumente empregado, como por Moritz *et al.* (2015), Klos, Bos e Bogusz (2018) e Nascimento *et al.* (2021).

6.2 Estimativa de velocidades

Inicialmente, visando verificar a melhor metodologia para obter as velocidades das estações a partir de suas séries temporais posicionais, analisou-se as velocidades estimadas

utilizando os seis métodos propostos. Salienta-se que as velocidades estimadas com o *minimum entropy* (ME) são disponibilizadas no SARI com menos casas decimais que os demais métodos, o que limitou a comparação.

Diante dos resultados, pode-se observar que a diferença entre as velocidades estimadas por regressão linear (RL), *Sklearn* e *Statsmodels* foi da ordem de 10^{-8} m/ano, e para o WLS, da ordem de 10^{-5} m/ano, ou seja, são diferenças desprezíveis para a finalidade de atualização de coordenadas, o que torna estes métodos intercambiáveis (como exemplificado com cinco estações na Tabela 1). De fato, Wang (2021) salienta que uma regressão linear simples é frequentemente utilizada para estimar a tendência linear a partir das séries temporais posicionais, mas que, apesar de diferentes abordagens serem empregadas na implementação dos mínimos quadrados, as velocidades estimadas são muito semelhantes para locais que experimentam movimentos lineares ao longo do tempo.

Tabela 1 - Velocidades estimadas com os diferentes métodos.

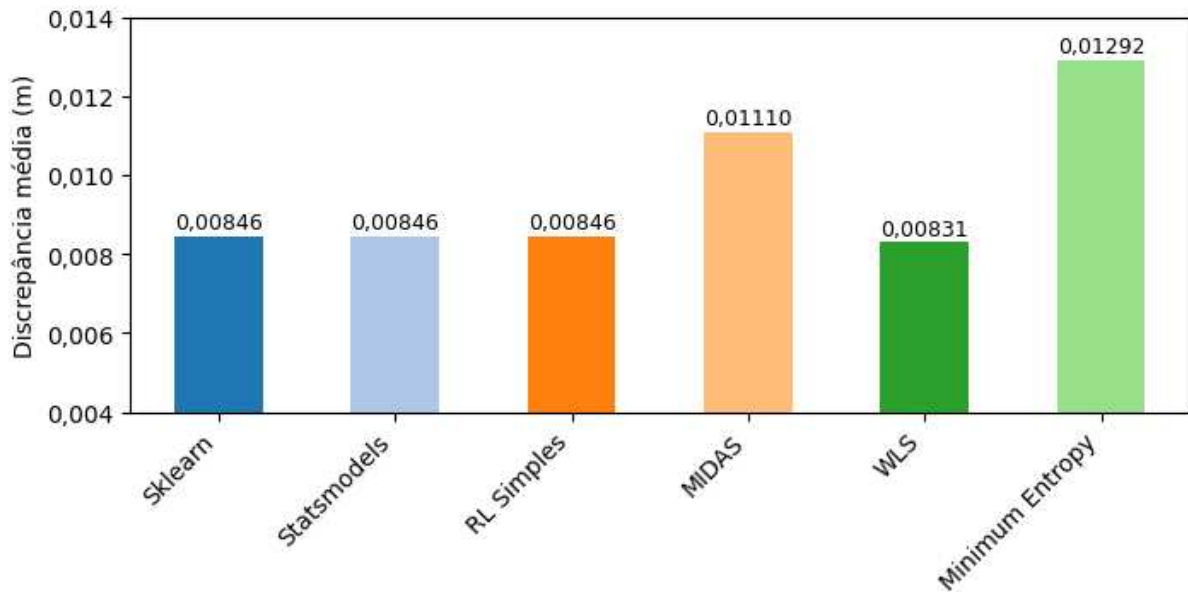
| Método | vx (m/ano) | vy (m/ano) | vz (m/ano) |
|--------------------|-------------------|-------------------|-------------------|
| AGMT | | | |
| <i>Sklearn</i> | -0,017578337590 | 0,006723661343 | -0,002704510749 |
| <i>Statsmodels</i> | -0,017578337569 | 0,006723661579 | -0,002704510870 |
| RL | -0,017578339641 | 0,006723646370 | -0,002704534311 |
| WLS | -0,017577791237 | 0,006722034740 | -0,002711260626 |
| MIDAS | -0,017419999930 | 0,006885000039 | -0,002390500158 |
| ME | -0,01758357 | 0,006733928 | -0,002969806 |
| CAND | | | |
| <i>Sklearn</i> | -0,018157614591 | 0,010000255492 | 0,001297070133 |
| <i>Statsmodels</i> | -0,018157614688 | 0,010000255208 | 0,001297070150 |
| RL | -0,018157605145 | 0,010000268107 | 0,001297091302 |
| WLS | -0,018160052592 | 0,010159427889 | 0,001484252720 |
| MIDAS | -0,019192500040 | 0,013899999670 | 0,005821999861 |
| ME | -0,01965246 | 0,01390640 | 0,006079311 |
| GARL | | | |
| <i>Sklearn</i> | -0,017595335481 | 0,004696317346 | -0,006045640440 |
| <i>Statsmodels</i> | -0,017595335442 | 0,004696317170 | -0,006045640394 |
| RL | -0,017595358342 | 0,004696298062 | -0,006045668800 |
| WLS | -0,017594628330 | 0,004697703263 | -0,006045748730 |
| MIDAS | -0,017512999940 | 0,004749000072 | -0,006196999922 |
| ME | -0,01760107 | 0,004706146 | -0,006036857 |
| NEWS | | | |
| <i>Sklearn</i> | -0,017713585880 | 0,002051458431 | -0,007043442489 |
| <i>Statsmodels</i> | -0,017713585886 | 0,002051458528 | -0,007043442145 |
| RL | -0,017713571753 | 0,002051455137 | -0,007043439643 |
| WLS | -0,017714939925 | 0,002050088761 | -0,007043271042 |
| MIDAS | -0,017598999660 | 0,002139889635 | -0,007048500003 |
| ME | -0,01770851 | 0,001933959 | -0,007014643 |
| OPCX | | | |
| <i>Sklearn</i> | -0,014620760133 | 0,002467729207 | -0,005324015412 |
| <i>Statsmodels</i> | -0,014620760247 | 0,002467729099 | -0,005324015393 |
| RL | -0,014620755590 | 0,002467711705 | -0,005324011095 |
| WLS | -0,014629852159 | 0,002465921326 | -0,005331475841 |
| MIDAS | -0,014399500100 | 0,002593000419 | -0,005029999651 |
| ME | -0,01491644 | 0,002486199 | -0,005860351 |

Fonte: elaboração própria.

Por outro lado, o MIDAS e o ME apresentaram diferenças na ordem de 10^{-3} m/ano em relação aos demais métodos, o que pode gerar diferenças mais significativas ao atualizar coordenadas. Na sequência, visando analisar o desempenho das velocidades sendo aplicadas no

processo de atualização, prosseguiu-se para o cálculo da discrepância planimétrica em relação às coordenadas de referência. Os resultados podem ser visualizados na Figura 19.

Figura 19 - Discrepância planimétrica média das 113 estações por método de estimativa de velocidade.



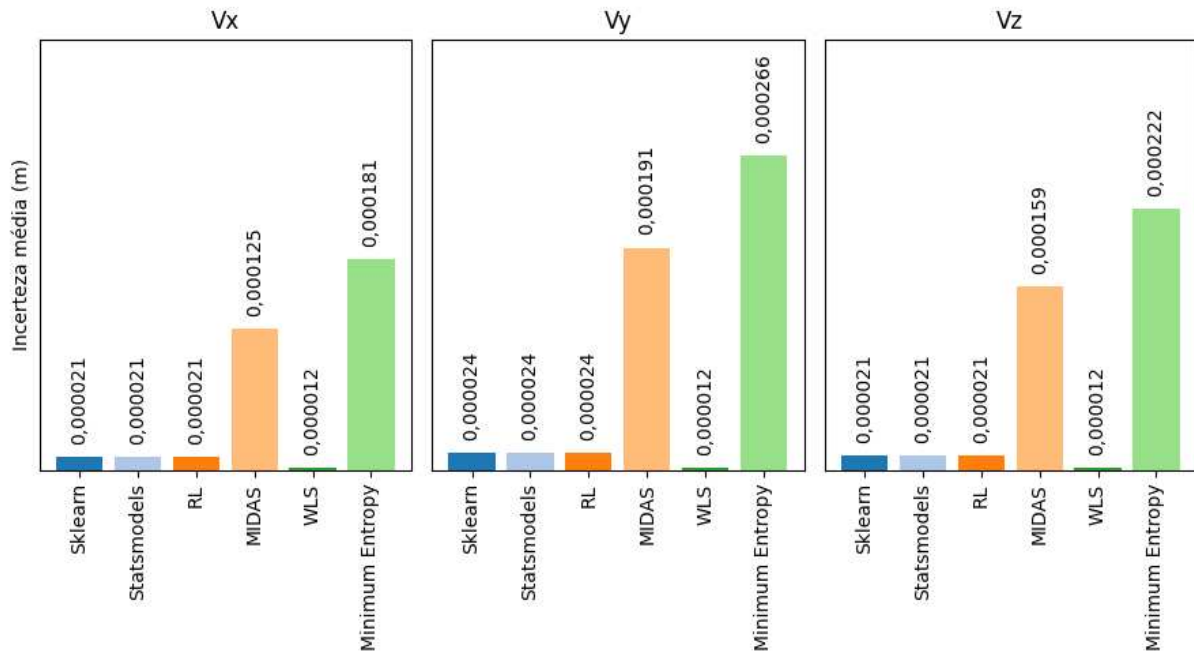
Fonte: elaboração própria.

Analisando a discrepância planimétrica média das 113 estações fica mais uma vez evidente que o MIDAS e o ME (discrepância de 0,011 m e 0,013 m, respectivamente) apresentaram uma diferença perceptível em relação aos demais métodos. Como os valores das velocidades estimadas por RL, *Sklearn* e *Statsmodels* eram similares, as discrepâncias médias relacionadas a estes métodos também apresentaram valores semelhantes (0,00846 m). Neste contexto, o WLS apresentou uma leve melhora (submilimétrica) com discrepância média de 0,00831 m. De fato, Greene (2018) e Wooldridge (2019) destacam que o WLS fornece estimativas mais eficientes e inferências mais precisas ao utilizar pesos que são inversamente proporcionais às variâncias dos resíduos, sendo o melhor estimador linear na presença de heterocedasticidade, isto é, quando a variância dos resíduos não é constante.

Com relação às incertezas das velocidades, os modelos MIDAS e ME - que são modelos robustos - apresentaram os maiores valores. Saleh, Bennett e Williams (2024) destacam que os dois métodos possuem incertezas bastante similares para séries curtas (~5 anos) e que com o aumento do período das séries há também um aumento na diferença entre as incertezas. A RL, *sklearn* e *statsmodels* apresentaram resultados idênticos, já o WLS foi o método com as menores incertezas (0,000012 m), como pode ser visualizado na Figura 20. Os gráficos das

incertezas individuais para as séries de cada estação estão disponíveis em https://bit.ly/incertezas_vel, e as tabelas estão disponíveis em https://bit.ly/tabelas_incertezas.

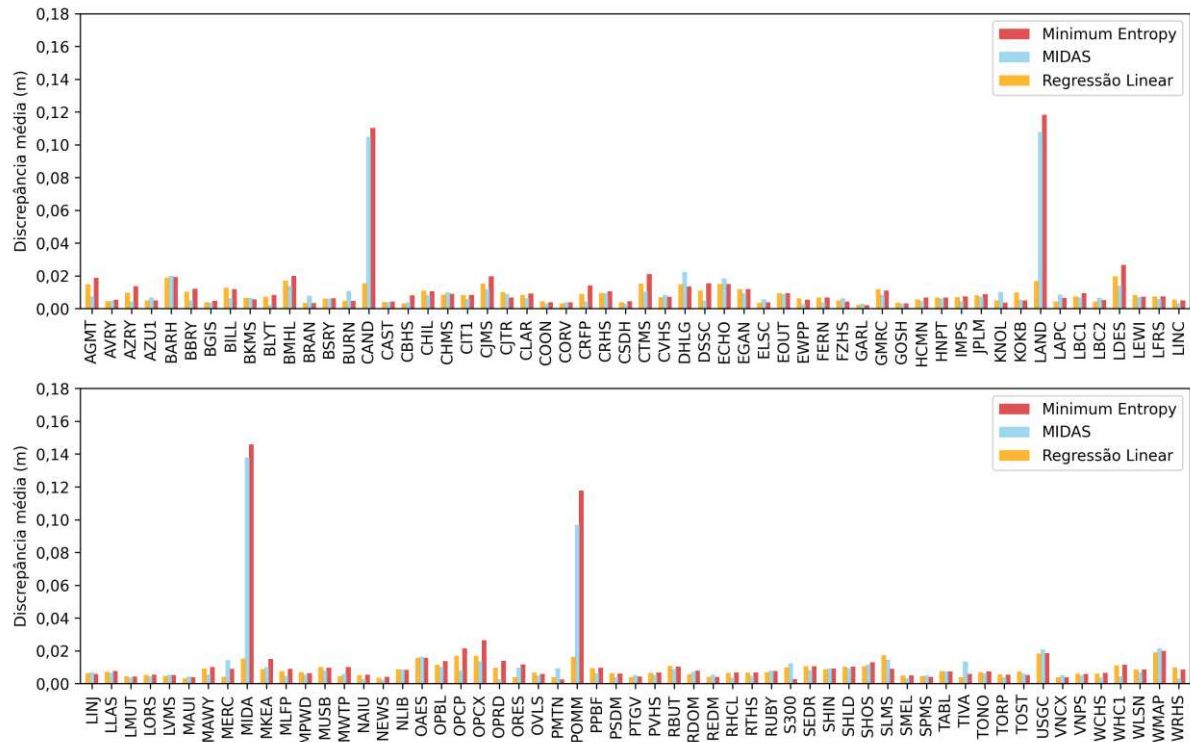
Figura 20 - Incertezas médias das velocidades das 113 estações por método de estimativa.



Fonte: elaboração própria.

Visando verificar as diferenças entre o MIDAS, o ME e a RL (que será doravante tratada como metodologia única, englobando os cinco métodos que apresentaram resultados semelhantes), inspecionou-se a discrepância para cada uma das 113 estações separadamente, como mostra a Figura 21.

Figura 21 - Discrepância planimétrica média para cada estação com o ME, MIDAS e regressão linear.

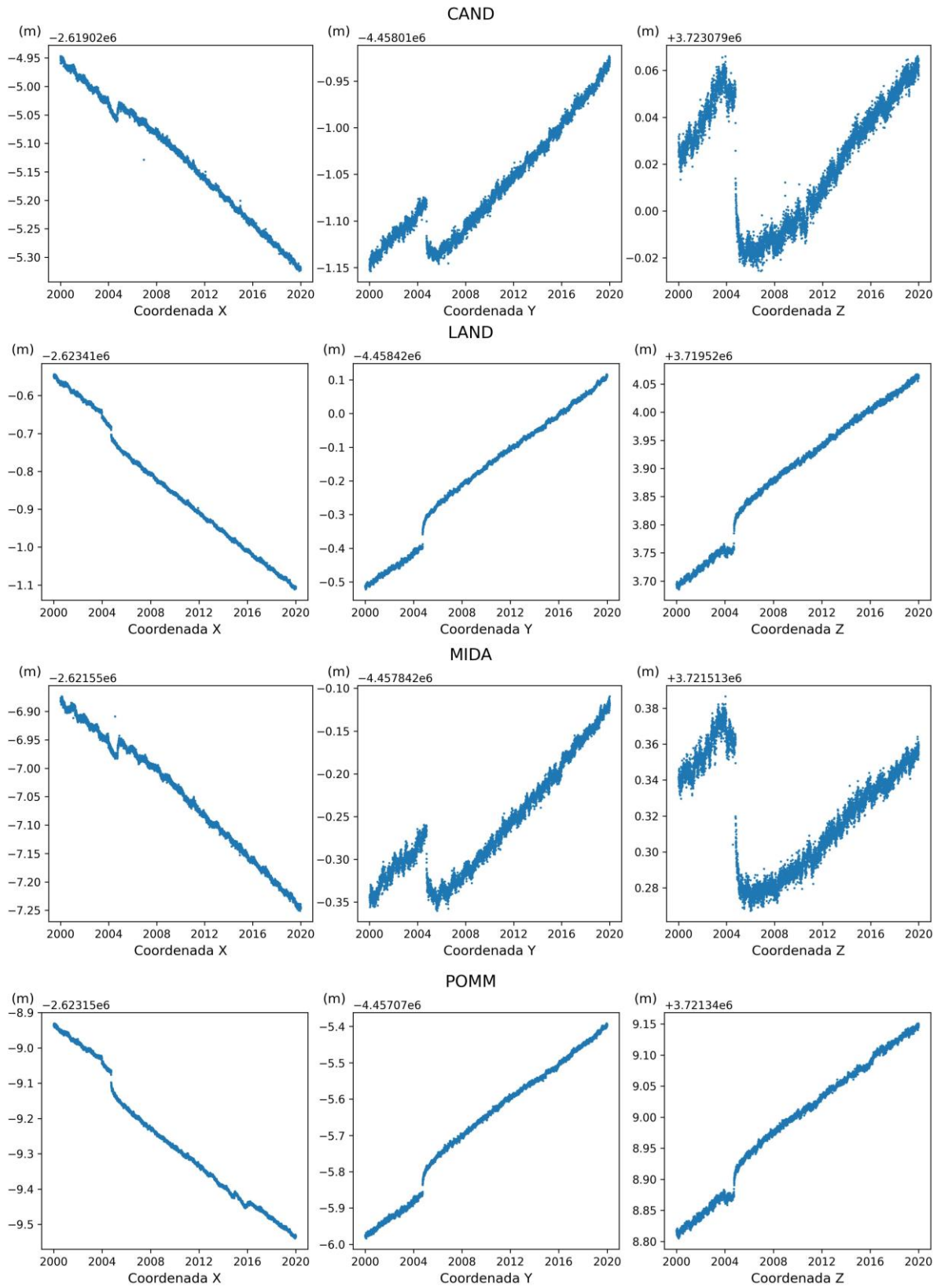


Fonte: elaboração própria.

Pelos resultados, nota-se uma diferença expressiva em quatro estações: CAND, LAND, MIDA e POMM, nas quais o MIDAS apresentou discrepâncias planimétricas de 0,105 m, 0,108 m, 0,138 m e 0,097 m, respectivamente; o ME de 0,110 m, 0,118 m, 0,146 m e 0,118 m, respectivamente; e a RL de 0,015 m, 0,017 m, 0,015 m e 0,016 m, respectivamente. Essa diferença é explicada pelas séries temporais das coordenadas destas quatro estações (Figura 22), que apresentam um salto significativo em 28 de setembro de 2004, causado por um terremoto de 6,0 na escala Richter com epicentro a 14,8 km da CAND, 13,3 km da LAND, 14,2 km da MIDA e 15,2 km da POMM. De fato, Blewitt et al. (2016) apontam que alterações no padrão de deslocamento das estações provocadas por atividades sísmicas podem comprometer a interpretação dos resultados do MIDAS, gerando valores que correspondem, por exemplo, às condições anteriores a tais atividades.

Em vista da similaridade dos resultados do MIDAS e do ME, vale destacar que ambos são métodos não paramétricos que quase não fazem exigências sobre a distribuição e estrutura de correlação das séries cujas velocidades serão estimadas; o que os faz, portanto, eficientes. No entanto, em todos os outros aspectos técnicos esses dois métodos são absolutamente distintos (Saleh; Bennett; Williams, 2024).

Figura 22 - Séries temporais posicionais das estações CAND, LAND, MIDA e POMM.



Fonte: elaboração própria.

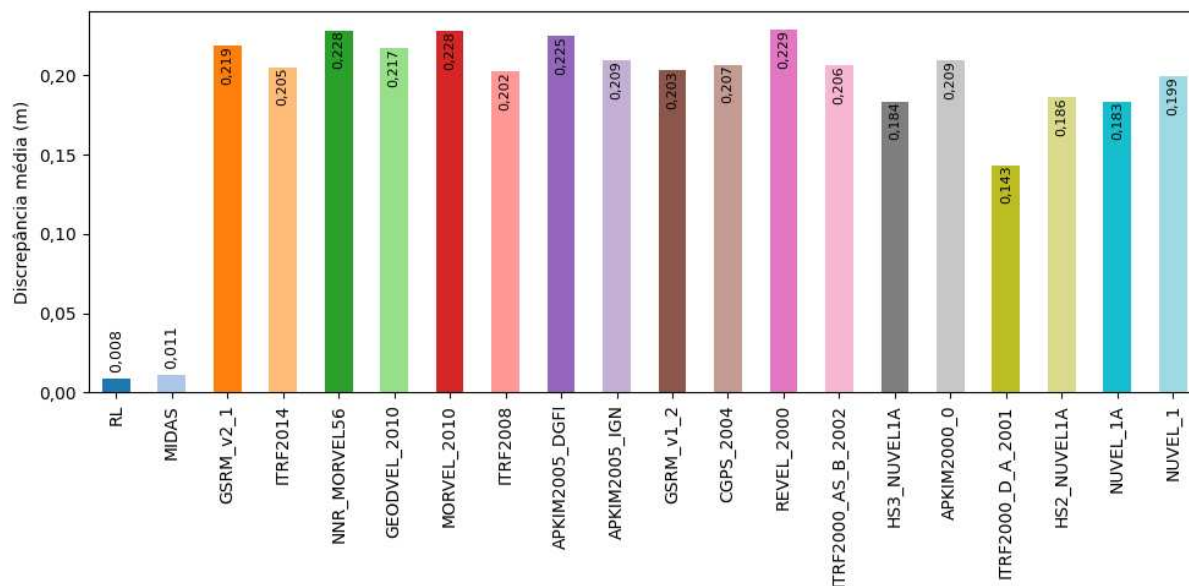
Neste contexto, comparado à RL, o MIDAS e o ME não apresentaram resultados satisfatórios nestas estações, não conseguindo estimar uma velocidade mais coerente devido aos saltos. Desconsiderando essas quatro estações, a discrepância média das demais 109 estações foi de 0,007 m para a RL, 0,008 m para o MIDAS e 0,009 para o ME. O que mostra que no geral as velocidades estimadas pelos três métodos são semelhantes, sendo apenas em alguns casos particulares que os resultados divergem. A discrepância planimétrica máxima com a RL entre todas as estações foi de 0,020 m, para a estação LDES, localizada em Homestead Valley, no estado da Califórnia, próximo ao encontro das Placas Norte-Americana e do Pacífico. De acordo com as informações da estação fornecidas pelo NGL, foram registrados três terremotos de magnitude entre 6,4 e 7,2 na escala Richter no período da série, com epicentro entre 187 e 244 km de distância da estação. Já com o MIDAS e o ME, a maior discrepância foi de 0,138 m e 0,146 m, respectivamente, na estação MIDA, já descrita acima. Em relação à menor discrepância registrada, com a RL e o ME foi de 0,002 m para a GARL, localizada em Washoe County School District, no estado de Nevada, noroeste dos EUA - um pouco mais afastado do limite da placa Norte-Americana. Para o MIDAS, o menor valor também foi de 0,002 m, para a NEWS, localizada entre New Pass Peak, também em Nevada.

Comparando os resultados dos três métodos de estimativa de velocidade para cada estação das 113, observou-se que o MIDAS apresentou menor discrepância planimétrica em 57,5% das estações, a RL em 27,4% e o ME 15,1%. Comparando apenas o MIDAS e a RL, o MIDAS apresentou menor discrepância em 62,83% das estações contra 37,17% da RL.

Desta forma, o MIDAS se mostrou um método mais indicado que o ME para estimativa de velocidades através de séries posicionais. Já em relação à RL, apesar da melhor performance relativa do MIDAS, em termos absolutos os valores não apresentaram diferenças significativas. Sendo assim, devido à maior facilidade de implementação e da maior consistência dos resultados, mesmo em séries com saltos, a RL deve ser o método preferido – e mais especificamente, a abordagem do WLS. Entretanto, vale destacar que as velocidades estimadas podem ser afetadas pela sazonalidade das séries, especialmente no caso de séries curtas (Blewitt; Lavallée, 2002; Blewitt *et al.*, 2016). Desta forma, o experimento ter sido realizado em séries de 20 anos pode ter contribuído para minimizar esta limitação da RL nos resultados.

Para avaliar as velocidades estimadas, comparou-se as discrepâncias planimétricas médias obtidas com aquelas de 18 modelos de velocidades disponibilizados pelo *Plate Motion Calculator* (Figura 23).

Figura 23 - Discrepância planimétrica média das 113 estações obtidas com as velocidades estimadas a partir das séries temporais posicionais e com 18 modelos de velocidades.



Fonte: elaboração própria.

Salienta-se que as velocidades fornecidas pela *Plate Motion Calculator* estão referenciadas ao WGS84 (*World Geodetic System 1984*). Burch (2016) destaca que o WGS84 e o ITRF2014 são alinhados e consistentes, não havendo parâmetros de transformação entre eles. Portanto, não houve necessidade de compatibilizar o referencial das velocidades antes do processo de atualização temporal.

Comparando as discrepâncias planimétricas fica clara a diferença de performance. Os modelos apresentaram discrepância média das 113 estações entre 0,143 m e 0,229 m, contra os já mencionados 0,008 m com a RL e 0,011 m com o MIDAS.

Dos 18 modelos utilizados, 6 são geofísicos (logo, levam em consideração a média de movimentação de centenas de milhares ou milhões de anos das placas litosféricas) e 12 possuem observações geodésicas, dos quais: cinco tem seu intervalo de dados observados anterior ao ano 2000, quatro cobrem entre três e sete anos do período de atualização estudado, e três cobrem entre 9,5 e 15 anos deste período.

O ITRF2000 (Drewes; Angermann, 2001) apresentou a menor discrepância entre os modelos, ainda que os dados utilizados para elaboração do modelo sejam referentes a um período anterior ao utilizado na atualização das coordenadas.

O modelo mais recente utilizado, o ITRF2014, que segundo Altamimi *et al.* (2016) utiliza dados de até 2015.1, abrangendo 75% do período de interesse, apresentou discrepância média de 0,205 m, consideravelmente pior que a RL ou o MIDAS. Dessa forma, quando o

período dos dados dos modelos ou da série temporal posicional utilizada para obter as velocidades e o período da atualização não são iguais, não é possível garantir que as velocidades sejam representativas da movimentação local, inserindo uma incerteza nos resultados. Logo, o modelo que irá apresentar o melhor resultado é quase aleatório.

Sánchez e Drewes (2016) apontam que eventos sísmicos significativos têm a capacidade de provocar alterações nos movimentos lineares das estações, resultando em mudanças nas suas velocidades. Da mesma forma, Castro, Ferreira e Freitas (2012) observaram variações na altitude geométrica de pontos distantes até 400 km do epicentro de um terremoto no Japão. Sendo assim, eventos locais podem modificar a tendência de longa duração, o que faz com que a velocidade pós-evento destoe da média de milhões de anos, como a utilizada pelos modelos geofísicos como o NUVEL 1, NUVEL 1A, HS2-NUVEL1A, HS3-NUVEL1A, MORVEL 2010 e NNR-MORVEL56. Com efeito, DeMets *et al.* (1994) identificaram que as Placa Norte-americana e do Pacífico estavam acelerando durante os anos 1990, com base em dados coletados no Golfo da Califórnia. Além disso, Argus *et al.* (2010) destacam que a velocidades das placas nas últimas décadas diferem significativamente das velocidades médias correspondentes dos últimos 3,2 milhões de anos.

Adicionalmente, diversos fatores locais como proximidade de grandes corpos d'água, topografia, tipo de solo etc. podem influenciar o deslocamento das estações. Em síntese, a soma desses fatores, aliado às incertezas inseridas pela interpolação das velocidades, explica as discrepâncias mais altas dos modelos.

6.3 Interpolação de velocidades

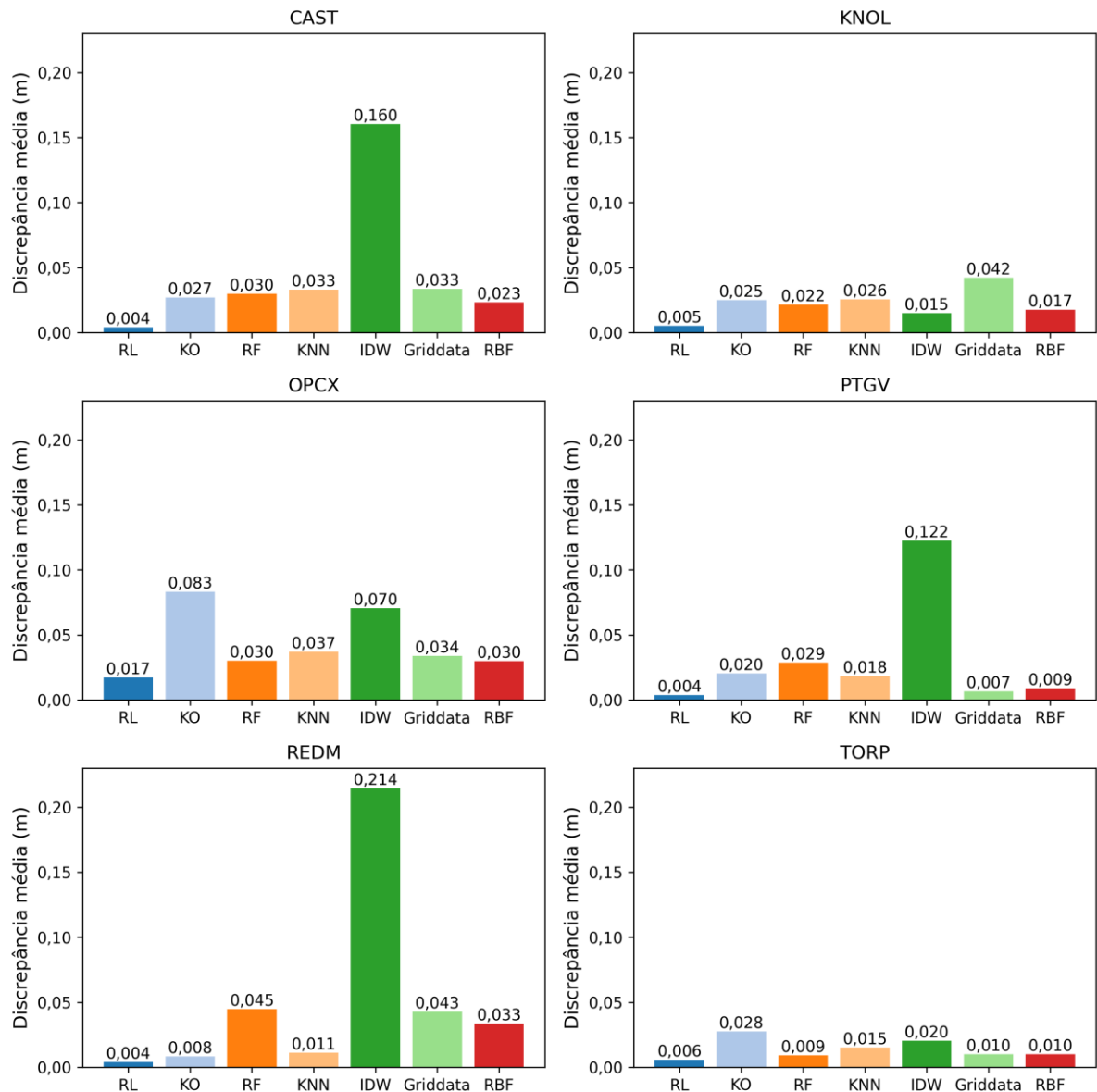
Uma vez de posse das velocidades das 113 estações, seguiu-se à interpolação para pontos desconhecidos, cujas disposições podem ser revisitadas na Figura 15. Destaca-se que não será analisada a diferença numérica entre os valores das velocidades estimadas através da série temporal e as interpoladas, uma vez que não é intuitivo o impacto que uma variação milimétrica ou submilimétrica implicará na utilização final da velocidade. Portanto, assim como nos outros experimentos a análise será baseada na discrepância planimétrica média calculada após a atualização de 365 coordenadas por 20 anos, e na comparação com os 18 modelos de velocidades.

As discrepâncias calculadas para as seis estações podem ser visualizadas na Figura 24. Para a estação CAST - em que as três estações vizinhas mais próximas estão a aproximadamente

160, 190 e 200 km - o *griddata*, KNN e RF apresentaram valores semelhantes entre 0,030 e 0,033 m; já a KO apresentou um valor levemente menor, de 0,027 m. A menor discrepância foi obtida com o RBF (0,023 m), o que representa 0,019 m acima do valor de referência da RL. Já o IDW apresentou a maior discrepância, com 0,160 m. A CAST está situada no estado de Utah, distante das bordas da placa Norte-Americana, entretanto, considerando que as estações vizinhas estavam todas em seu lado ocidental, não sendo uma distribuição ótima para a interpolação, os resultados possuem qualidade suficiente para serem compatíveis com diversas aplicações práticas.

A estação KNOL, localizada no norte da Califórnia, apresenta as estações vizinhas mais próximas dentre as seis selecionadas (localizadas a 3,5 km, 4 km e 7,5 km). A maior discrepância foi registrada com o método *Griddata* (0,042 m), enquanto os demais métodos apresentaram valores que variam entre 0,015 m e 0,026 m. Esta estação foi a única em que o IDW foi o melhor método, isto reflete o maior peso atribuído pelo IDW às estações mais próximas na interpolação, fazendo com que o valor interpolado seja similar aos conhecidos. Ademais, considerando as características locais, a proximidade também faz com que não haja uma variação na movimentação crustal em uma extensão de menos de 5 km.

Figura 24 - Discrepância planimétrica média por método de interpolação para cada estação.



Fonte: elaboração própria.

A OPCX, situada no sul da Califórnia, apesar de estar mais próxima de suas estações vizinhas (9,5, 17,5 e 22 km), apresentou resultados similares àqueles da CAST, com os métodos *griddata*, KNN, RF e RBF possuindo valores no intervalo de 0,030 a 0,037 m. Os métodos com maiores discrepâncias foram a KO e o IDW: 0,083 e 0,070 m, respectivamente. Destaca-se que a discrepância de referência já era de um valor mais alto, única estação acima de 1 cm (0,017 m). Foram registrados três terremotos durante o período da série entre 6,4 e 7,2 na escala Richter, com epicentro entre 290 e 250 km de distância da estação.

A PTGV, apesar de ser a estação com as estações vizinhas mais distantes (290, 620 e 1200 km), apresentou as menores discrepâncias dentre as seis estações estudadas. Com o *griddata* apresentando uma discrepância (0,007 m) apenas 1 mm acima do valor de referência (0,006 m), e o RBF apenas 3 mm acima (0,09 m). Os valores para o KNN, KO e RF foram de 0,018 m, 0,020 m e 0,029 m, respectivamente. A baixa discrepância pode ser atribuída ao fato de a estação estar localizada no leste dos EUA, região central da placa Norte-Americana, e conseqüentemente, mais estável, o que pode ser observado pelo seu vetor velocidade $\vec{v}_{m/ano} = (-0,014, 0,001, -0,001)$, com módulo de 1,4 cm/ano.

Destaca-se que as atividades sísmicas impactam as estações de forma diferente, dependendo, por exemplo, da topografia e tipo de solo da região e da proximidade dos epicentros dos terremotos. Desta forma, se uma estação utilizada como ponto conhecido na interpolação possuir saltos e/ou falhas, isso irá afetar a precisão de sua velocidade, o que conseqüentemente afetará a velocidade interpolada.

A REDM, localizada no estado de Oregon, possui suas três estações vizinhas mais próximas localizadas a 175, 315 e 320 km. O IDW apresentou a maior discrepância entre os resultados de todas as estações, com 0,214 m. A KO e o KNN, por outro lado, apresentaram a menor discrepância entre as 6 estações (0,008 e 0,011 m, respectivamente), o que equivale a 4 e 7 mm acima do valor de referência, respectivamente. Os demais métodos apresentaram discrepâncias entre 0,033 e 0,045 m.

Entre as estações escolhidas, a TORP está localizada na região de maior atividade sísmica. Assim como a KNOL OPCX, ela também está situada no estado da Califórnia, mas na placa do Pacífico - próxima da borda e a menos de 10km do litoral. A influência de tais características na posição da estação pode ser observada em seu vetor velocidade $\vec{v}_{m/ano} = (-0,0300, 0,028, 0,016)$, com módulo de mais de 4 cm/ano. As três estações vizinhas mais próximas estão a 4,5, 6 e 10 km. Apesar da região, as discrepâncias foram baixas, com o melhor métodos sendo o RF, com 0,009 m (3 mm acima do valor de referência). O *griddata* e o RBF tiveram o mesmo valor de 0,010 m e o KNN, IDW e KO, 0,015 m, 0,020 m, e 0,028 m respectivamente.

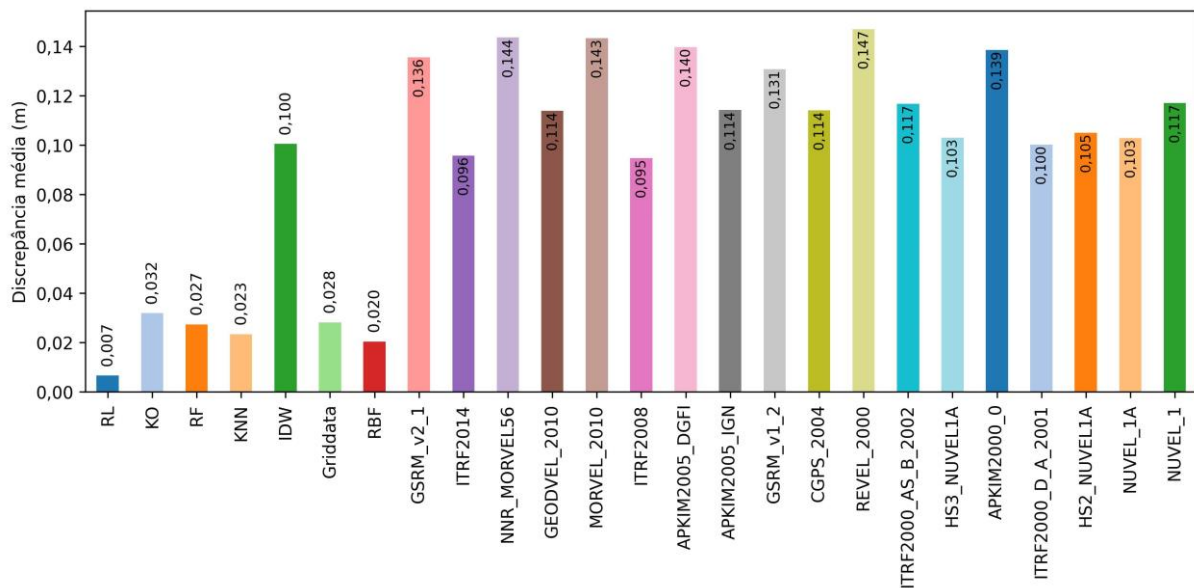
De modo geral, não foi possível observar características específicas dos métodos de modo a atribuir um melhor resultado a uma dada configuração geográfica das estações vizinhas. O IDW teve o melhor resultado na KNOL, mas foi o pior em outras 3 estações. O *griddata* foi o melhor para a estação PTGV, mas foi o pior na KNOL. A KO foi o melhor na OPCX e a pior

na REDM e TORP. O RF, o KNN e o RBF se mostraram mais consistentes, com resultados satisfatórios em todas as estações.

Calculando a discrepância média das seis estações para cada método, verifica-se que o IDW com 0,100 m foi o de pior desempenho. A KO, *griddata* e RF foram bem similares, com 0,032 m, 0,028 m e 0,027 m, respectivamente. Os dois melhores foram o KNN com 0,023 m e o RBF com 0,020 m. Devido à maior consistência entre todas as estações, o RBF se mostrou o método mais confiável para interpolar velocidades estimadas a partir de séries temporais posicionais de observações GPS. Vale destacar que os experimentos foram realizados utilizando mais de 100 estações como pontos conhecidos, realizar interpolações utilizando apenas os três pontos mais próximos, por exemplo, pode apresentar resultados distintos, devido às particularidades de cada método.

A comparação desses resultados com as discrepâncias calculadas utilizando as velocidades fornecidas por 18 modelos de velocidades está disposta na Figura 25.

Figura 25 - Discrepância planimétrica média das seis estações consideradas desconhecidas: velocidades interpoladas e 18 modelos de velocidades.



Fonte: elaboração própria.

A discrepância média entre os 18 modelos variou entre 0,095 e 0,147 m, significativamente maior que o RF, KNN, *griddata*, KO e RBF. Entre os métodos de interpolação, somente o IDW apresentou resultado pior que algum dos modelos.

Tomando como referência o valor da RL, os métodos de interpolação RF, KNN, *griddata* e RBF apresentaram discrepâncias aproximadamente 3 a 4 vezes maiores. Contudo, as discrepâncias calculadas utilizando as velocidades interpoladas com estes métodos foram cerca de 3,5 a 7 vezes menores que as calculadas utilizando as velocidades dos 18 modelos analisados.

6.4 Identificação e remoção de *outliers*

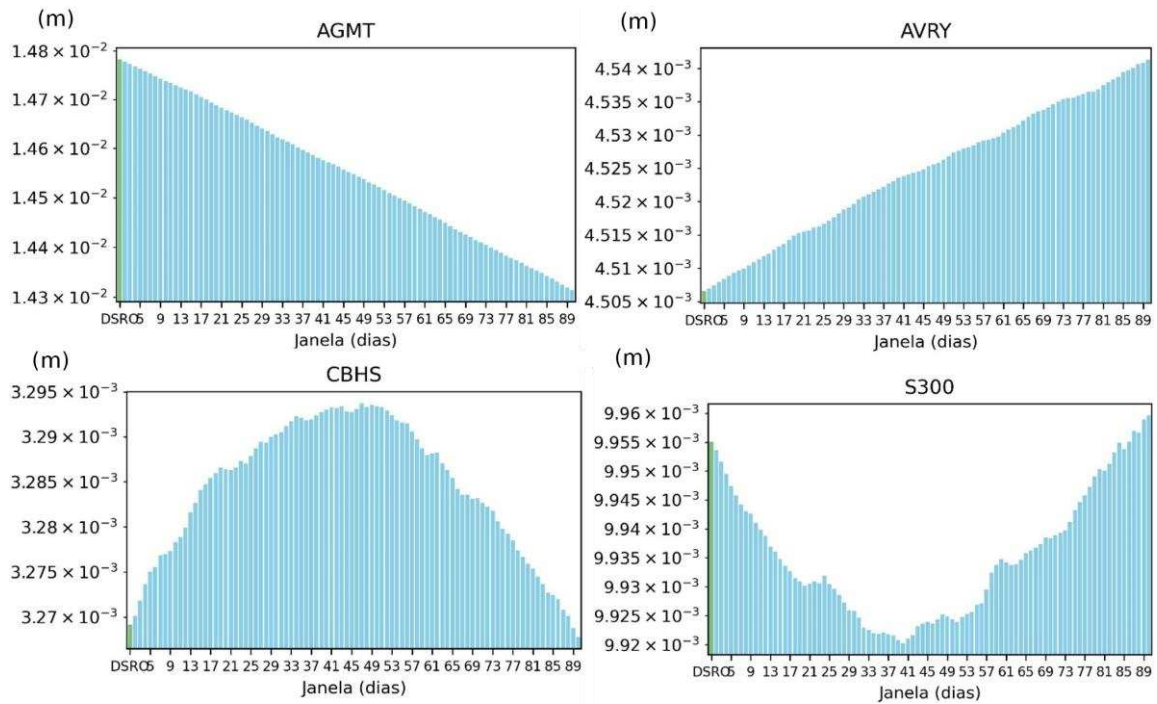
Como explicado na metodologia, os métodos de identificação de *outliers* foram aplicados nas séries após a remoção de tendência, com exceção da média móvel simples e do BCP-HI, que não exigem esse pré-requisito dos dados.

Destaca-se também que as tabelas presentes em todos os subtópicos dos métodos - por uma questão de espaço - compreendem uma lista idêntica de 29 das 113 estações. A seleção foi realizada de modo a compilar as estações que apresentaram resultados mais significativos em qualquer um dos métodos. As tabelas completas podem ser consultadas em bit.ly/Rem_Tend.

6.4.1 Média móvel simples

Após a aplicação das médias móveis, foram identificados quatro padrões no comportamento das discrepâncias para janelas de tamanho entre 2 e 90: i) de forma praticamente linear, quanto maior o tamanho da janela, menor a discrepância; ii) de forma praticamente linear, quanto maior a janela, maior a discrepância; iii) a discrepância aumenta com o aumento do tamanho da janela até um ponto de inflexão; e, iv) a discrepância diminui com o aumento do tamanho da janela até um ponto de inflexão. Esses comportamentos são exemplificados na Figura 26. Destaca-se que a escala dos gráficos foi exagerada (logarítmica) para melhorar a visualização dos comportamentos, mas a diferença entre a maior e a menor discrepância na maioria dos casos é submilimétrica.

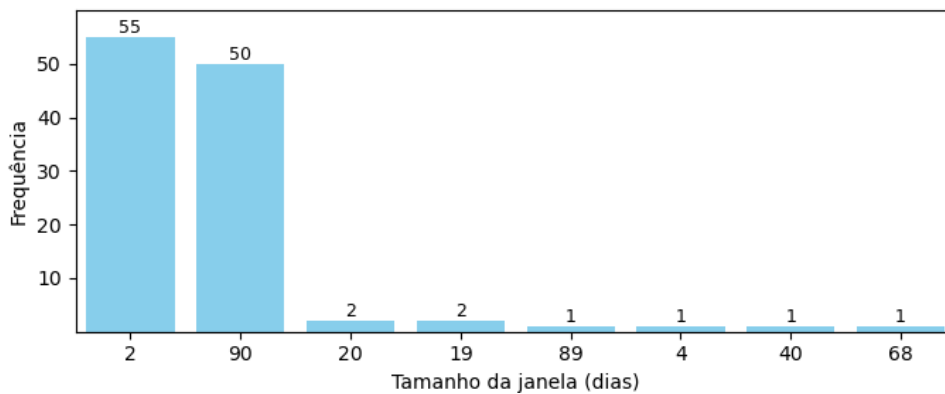
Figura 26 - Discrepância Sem Remoção de Outliers (DSRO) e discrepâncias após aplicação de média móvel para as estações AGMT, AVRY, CBHS e S300.



Fonte: elaboração própria.

Ao analisar os tamanhos de janela que proporcionaram as menores discrepâncias para cada estação, observa-se que, em 55 das 113 estações (48,7%), o melhor desempenho foi alcançado com janelas de tamanho 2. Outras 50 estações (44,2%) apresentaram os melhores resultados com janelas de tamanho 90, enquanto apenas 8 estações indicaram preferências por outros tamanhos, conforme ilustrado na Figura 27.

Figura 27 - Frequência de tamanho de janela que proporcionou a menor discrepância nas 113 estações.



Fonte: elaboração própria.

Após a aplicação da média móvel, a redução média da discrepância foi de 0,000113 m. A redução mais significativa foi de 0,000835 m, para a estação OAES. E a discrepância aumentou em 54 das 113 estações (47,8%). A Tabela 2 expõe os valores de discrepância calculados após a aplicação das médias móveis, antes da remoção dos *outliers* (Discrepância sem remoção de *outliers* - DSRO), bem como a diferença entre esses valores (Δ Discrepância), para o conjunto selecionado de estações.

Tabela 2 - Discrepâncias antes e depois da aplicação de média móvel simples.

| Estação | Tamanho da janela | Discrepância com média móvel (m) | DSRO (m) | Δ Discrepância (m) |
|----------------|--------------------------|---|-----------------|---|
| BARH | 90 | 0,014314 | 0,014783 | 0,000469 |
| BURN | 90 | 0,014314 | 0,014783 | 0,000469 |
| CAND | 2 | 0,004507 | 0,004507 | 0,000000 |
| CAST | 90 | 0,014314 | 0,014783 | 0,000469 |
| CJMS | 90 | 0,014314 | 0,014783 | 0,000469 |
| CLAR | 2 | 0,004507 | 0,004507 | 0,000000 |
| CRHS | 2 | 0,004507 | 0,004507 | 0,000000 |
| KNOL | 90 | 0,014314 | 0,014783 | 0,000469 |
| KOKB | 90 | 0,014314 | 0,014783 | 0,000469 |
| LAND | 2 | 0,004507 | 0,004507 | 0,000000 |
| LMUT | 2 | 0,004507 | 0,004507 | 0,000000 |
| MAUI | 90 | 0,014314 | 0,014783 | 0,000469 |
| MIDA | 2 | 0,004507 | 0,004507 | 0,000000 |
| OAES | 90 | 0,014314 | 0,014783 | 0,000469 |
| OPBL | 90 | 0,014314 | 0,014783 | 0,000469 |
| OPCP | 90 | 0,014314 | 0,014783 | 0,000469 |
| ORES | 90 | 0,014314 | 0,014783 | 0,000469 |
| OVLS | 2 | 0,004507 | 0,004507 | 0,000000 |
| PMTN | 90 | 0,014314 | 0,014783 | 0,000469 |
| POMM | 2 | 0,004507 | 0,004507 | 0,000000 |
| PPBF | 90 | 0,014314 | 0,014783 | 0,000469 |
| PSDM | 2 | 0,004507 | 0,004507 | 0,000000 |
| PVHS | 2 | 0,004507 | 0,004507 | 0,000000 |
| S300 | 40 | 0,009920 | 0,009955 | 0,000035 |
| SHIN | 20 | 0,008820 | 0,008828 | 0,000007 |
| TIVA | 90 | 0,014314 | 0,014783 | 0,000469 |
| TOST | 19 | 0,003982 | 0,003986 | 0,000004 |
| WLSN | 2 | 0,004507 | 0,004507 | 0,000000 |
| WMAP | 68 | 0,018965 | 0,019050 | 0,000085 |

Fonte: elaboração própria.

Em vista destes resultados, fica claro que a aplicação de médias móveis em séries temporais posicionais longas não afeta a estimativa de velocidade e conseqüentemente a discrepância planimétrica.

6.4.2 *Boxplot Ajustado*

Após a remoção da tendência das séries e a aplicação do *boxplot* ajustado, observou-se uma redução média na discrepância de 0,000320 m. A maior redução registrada foi de 0,006795 m, verificada na estação S300. Por outro lado, em 60 estações, houve um aumento médio na discrepância de 0,000289 m.

A Tabela 3 mostra a discrepância antes e depois da aplicação do *boxplot* ajustado, qual *threshold* proporcionou a menor discrepância e a contaminação das séries para algumas das estações.

Tabela 3 - Discrepâncias antes e após remoção de *outliers* com *boxplot* ajustado.

| Estação | Threshold | Discrepância Boxplot Aj. (m) | DSRO (m) | Δ Discrepância (m) | Contaminação (%) |
|---------|-----------|------------------------------------|-------------|------------------------------|---------------------|
| BARH | 2 | 0,018744 | 0,018779 | 0,000035 | 1,57 |
| BURN | 4,5 | 0,004447 | 0,004641 | 0,000194 | 2,01 |
| CAND | 5 | 0,015373 | 0,015411 | 0,000038 | 1,38 |
| CAST | 4 | 0,003994 | 0,004006 | 0,000012 | 1,22 |
| CJMS | 2 | 0,017775 | 0,015322 | -0,002453 | 3,34 |
| CLAR | 3,5 | 0,008246 | 0,008226 | -0,000019 | 2,43 |
| CRHS | 2 | 0,010597 | 0,009449 | -0,001147 | 5,23 |
| KNOL | 2 | 0,004932 | 0,005065 | 0,000133 | 1,85 |
| KOKB | 2 | 0,007111 | 0,009953 | 0,002842 | 7,83 |
| LAND | 5 | 0,016220 | 0,016858 | 0,000638 | 1,59 |
| LMUT | 2 | 0,004937 | 0,004498 | -0,000439 | 1,75 |
| MAUI | 5 | 0,003315 | 0,003269 | -0,000046 | 1,83 |
| MIDA | 5 | 0,015154 | 0,015183 | 0,000029 | 0,04 |
| OAES | 2 | 0,015754 | 0,015666 | -0,000088 | 4,67 |
| OPBL | 5 | 0,011629 | 0,011413 | -0,000216 | 1,24 |
| OPCP | 5 | 0,016812 | 0,017045 | 0,000233 | 3,13 |
| ORES | 5 | 0,003787 | 0,003802 | 0,000015 | 0,09 |
| OVLS | 3 | 0,006825 | 0,006808 | -0,000017 | 1,41 |
| PMTN | 4 | 0,003697 | 0,003848 | 0,000150 | 1,51 |
| POMM | 5 | 0,015124 | 0,016353 | 0,001229 | 4,01 |
| PPBF | 5 | 0,009595 | 0,009479 | -0,000117 | 1,56 |
| PSDM | 2 | 0,006274 | 0,006222 | -0,000052 | 1,76 |

| | | | | | |
|------|---|----------|----------|-----------------|------|
| PVHS | 2 | 0,006852 | 0,006636 | -0,000216 | 2,19 |
| S300 | 2 | 0,003161 | 0,009955 | 0,006795 | 1,36 |
| SHIN | 5 | 0,008823 | 0,008762 | -0,000060 | 1,24 |
| TIVA | 5 | 0,003900 | 0,003954 | 0,000054 | 6,91 |
| TOST | 5 | 0,005551 | 0,007375 | 0,001824 | 2,13 |
| WLSN | 2 | 0,009007 | 0,008562 | -0,000445 | 2,79 |
| WMAP | 2 | 0,018424 | 0,019050 | 0,000626 | 2,93 |

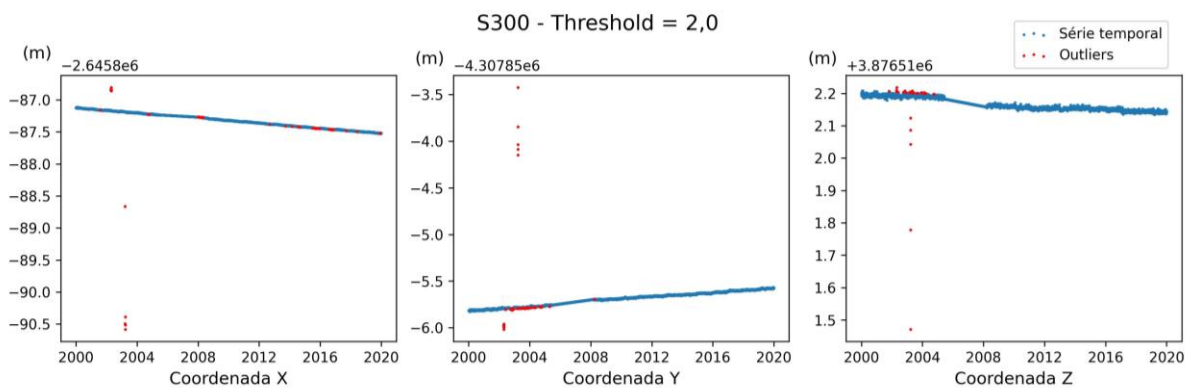
Fonte: elaboração própria.

Das 113 estações, apenas quatro apresentaram redução de mais de 1 mm, sendo elas: KOKB, POMM, S300 e TOST.

A KOKB passou por uma troca de antena e receptor em 2002 e uma nova troca de antena em 2004, que ocasionaram um salto em sua série. A TOST passou por uma troca de antena em setembro de 2000 que também provocou um salto na série. A WMAP apresenta um pequeno salto em 2004, entretanto, não foram registrados nenhum evento de troca de equipamentos ou terremotos para o período.

A S300 apresenta cinco *outliers* de grande magnitude no primeiro semestre de 2003, com o pior chegando a um deslocamento em relação à linha de tendência da série que passa de 3 m na coordenada X, 2 m na coordenada Y e de 0,60 m na coordenada Z, como mostra a Figura 28. Não foram registrados nenhum evento de troca de equipamento ou terremotos para o período.

Figura 28 - *Outliers* identificados com *boxplot* ajustado para a estação S300.

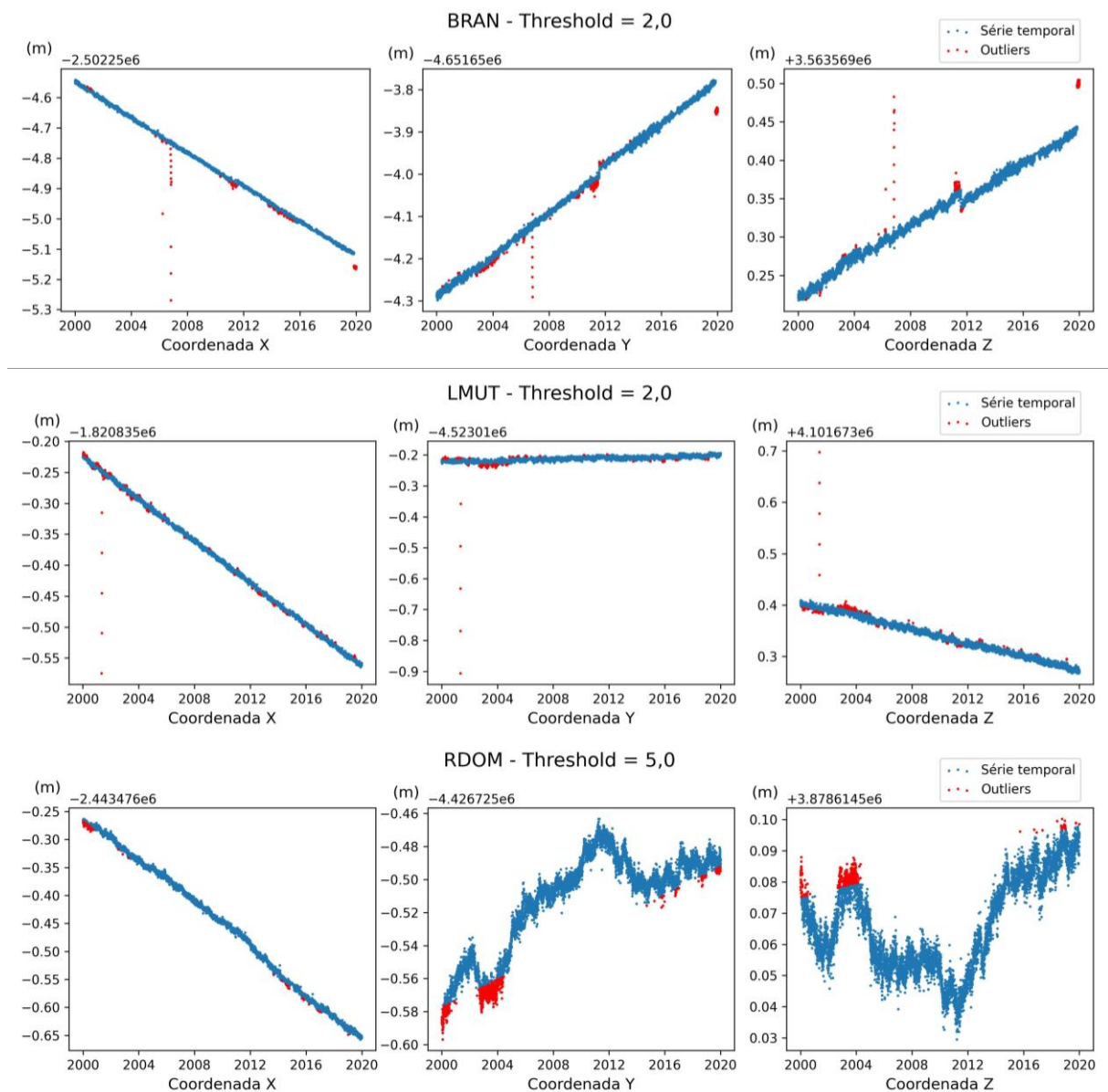


Fonte: elaboração própria.

Considerando as 113 estações analisadas, a contaminação média foi de 2,16%. O valor máximo foi de 7,83%, registrado na estação KOKB, enquanto o valor mínimo (0,04%) foi observado na estação MIDA.

O *boxplot* ajustado apresentou resultados bem satisfatórios, identificando uma alta porcentagem dos pontos que em uma identificação visual seriam classificados indubitavelmente como *outliers*, como ilustrado na Figura 29. Entretanto, em estações com movimentações menos lineares, como a RDOM, este método encontrou maior dificuldade para identificar os *outliers* ao longo da série; identificando, por outro lado, trechos com múltiplos pontos que se afastaram da tendência de longa duração.

Figura 29 - *Outliers* identificados com o *boxplot* ajustado nas estações BRAN, LMUT e RDOM.

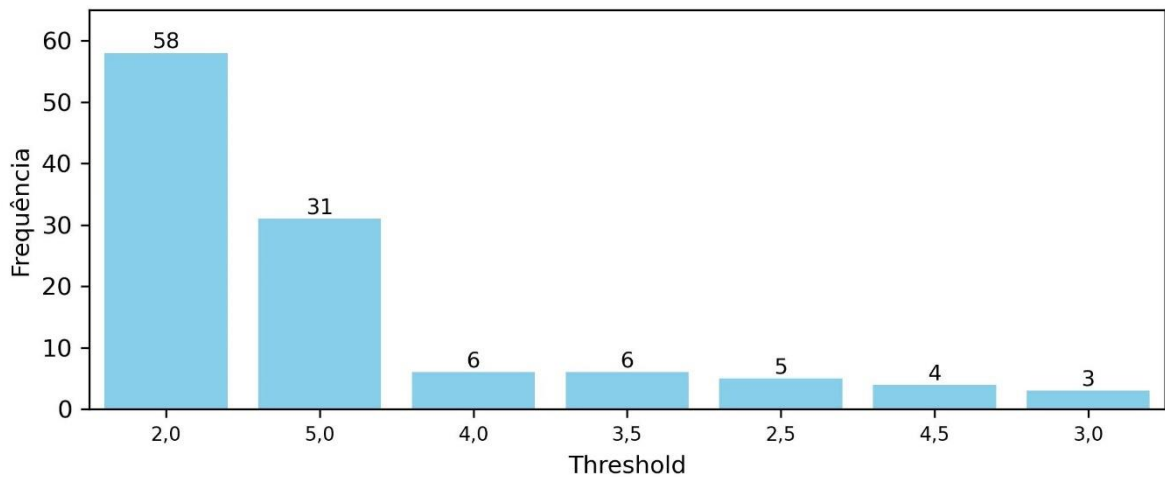


Fonte: elaboração própria.

Outro ponto a se destacar é que em estações como a BRAN e LMUT, que possuem *outliers* grosseiros como a S300, apresentaram um aumento na discrepância após a remoção dos *outliers*: 0,000032 m e 0,000439 m, respectivamente. Isto pode indicar uma baixa influência dos *outliers* no valor da velocidade estimada.

O *threshold* equivalente a 2 proporcionou a menor discrepância em 58 estações (51,3% das 113 estações analisadas), seguido pelo *threshold* igual a 5 com 31 estações (27,4% das estações). Os demais valores somados totalizaram 24 estações (Figura 30).

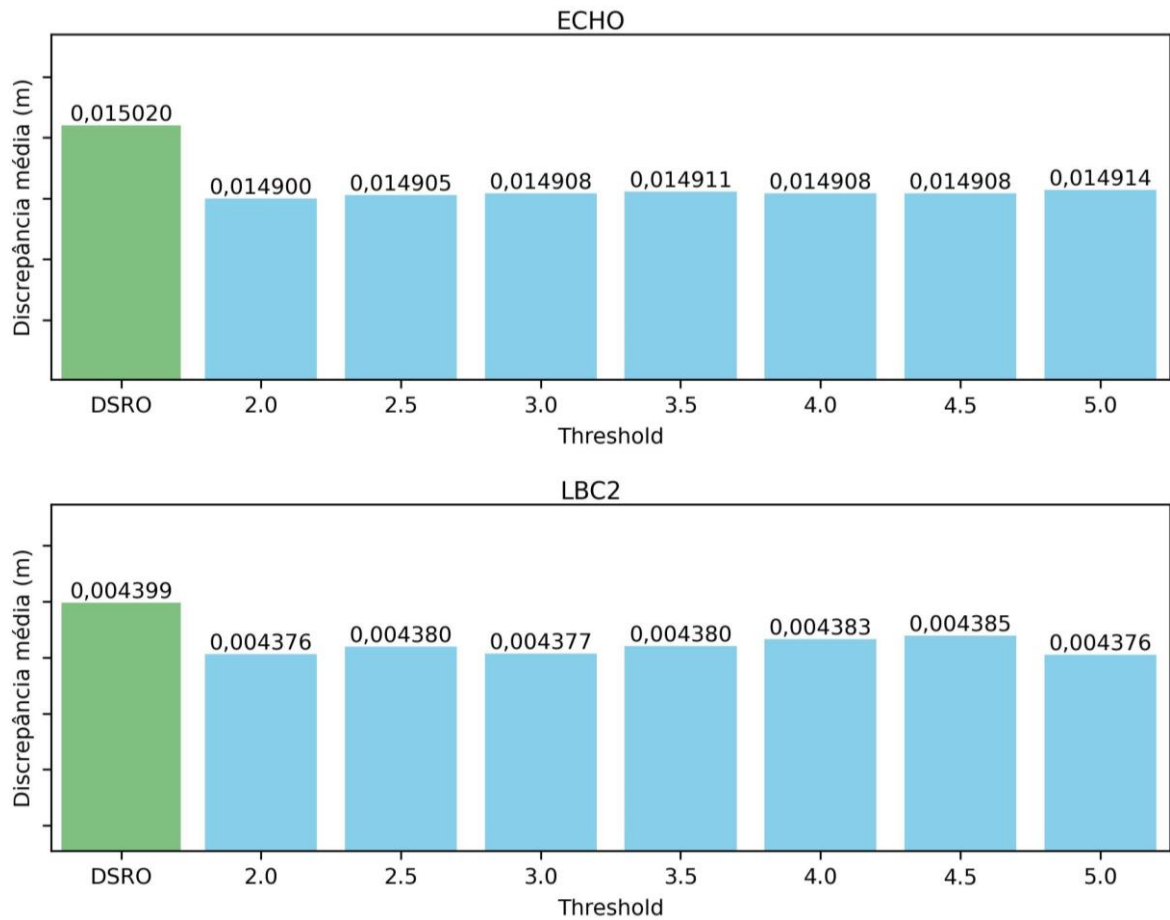
Figura 30 - Frequência do *threshold* que proporcionou a menor discrepância para as 113 estações, após aplicação do *boxplot* ajustado.



Fonte: elaboração própria.

Na maioria das estações a variação do *threshold* não influenciou significativamente a discrepância, como pode ser observado nas estações ECHO e LBC2 (Figura 31)

Figura 31 - Discrepâncias para cada valor de *threshold* nas estações ECHO e LBC2.



Fonte: elaboração própria.

O *threshold* normalmente empregado neste método é de 3 ou 3,5, como utilizado por Vandervieren e Hubert (2004), Hubert e Vandervieren (2008) e Ferreira *et al.* (2019), e que neste experimento não apresentaram os melhores resultados.

6.4.3 Z-Score

Após a remoção da tendência das séries e a aplicação do *Z-Score*, observou-se uma redução média na discrepância de 0,000256 m. A maior redução registrada foi de 0,007753 m, verificada na estação S300. Por outro lado, em 24 estações, houve um aumento médio na discrepância de 0,000144 m.

A Tabela 4 apresenta a discrepância antes e depois da aplicação do *Z-Score*, qual *Z-Score* proporcionou a menor discrepância e a contaminação das séries para algumas das estações.

Tabela 4 - Discrepâncias antes e após eliminação de *outliers* com *Z-Score*.

| Estação | <i>Z-Score</i> | Discrepância <i>Z-Score</i> (m) | DSRO (m) | Δ Discrepância (m) | Contaminação (%) |
|---------|----------------|------------------------------------|-------------|------------------------------|---------------------|
| BARH | 4,5 | 0,018758 | 0,018779 | 0,000020 | 0,03 |
| BURN | 2 | 0,004344 | 0,004641 | 0,000298 | 5,24 |
| CAND | 4 | 0,015416 | 0,015411 | -0,000004 | 0,00 |
| CAST | 2 | 0,003981 | 0,004006 | 0,000026 | 4,83 |
| CJMS | 5 | 0,015729 | 0,015322 | -0,000407 | 0,44 |
| CLAR | 3 | 0,008233 | 0,008226 | -0,000006 | 0,73 |
| CRHS | 5 | 0,009493 | 0,009449 | -0,000044 | 0,24 |
| KNOL | 2 | 0,004729 | 0,005065 | 0,000336 | 1,67 |
| KOKB | 2 | 0,007121 | 0,009953 | 0,002831 | 6,98 |
| LAND | 2 | 0,016755 | 0,016858 | 0,000103 | 0,65 |
| LMUT | 2 | 0,004958 | 0,004498 | -0,000459 | 0,18 |
| MAUI | 5 | 0,003327 | 0,003269 | -0,000057 | 0,04 |
| MIDA | 3 | 0,015175 | 0,015183 | 0,000008 | 0,00 |
| OAES | 2 | 0,015344 | 0,015666 | 0,000322 | 5,54 |
| OPBL | 4,5 | 0,011413 | 0,011413 | 0,000000 | 0,00 |
| OPCP | 5 | 0,017045 | 0,017045 | 0,000000 | 0,00 |
| ORES | 2 | 0,003793 | 0,003802 | 0,000008 | 0,72 |
| OVLS | 4,5 | 0,006802 | 0,006808 | 0,000006 | 0,01 |
| PMTN | 5 | 0,003526 | 0,003848 | 0,000322 | 0,33 |
| POMM | 2 | 0,016250 | 0,016353 | 0,000103 | 0,69 |
| PPBF | 5 | 0,009479 | 0,009479 | 0,000000 | 0,00 |
| PSDM | 5 | 0,006219 | 0,006222 | 0,000004 | 0,35 |
| PVHS | 2 | 0,006590 | 0,006636 | 0,000046 | 3,75 |
| S300 | 5 | 0,002202 | 0,009955 | 0,007753 | 0,07 |
| SHIN | 5 | 0,008762 | 0,008762 | 0,000000 | 0,00 |
| TIVA | 2 | 0,003859 | 0,003954 | 0,000095 | 5,34 |
| TOST | 2 | 0,005948 | 0,007375 | 0,001426 | 2,69 |
| WLSN | 5 | 0,008690 | 0,008562 | -0,000127 | 0,57 |
| WMAP | 2 | 0,017953 | 0,019050 | 0,001097 | 4,66 |

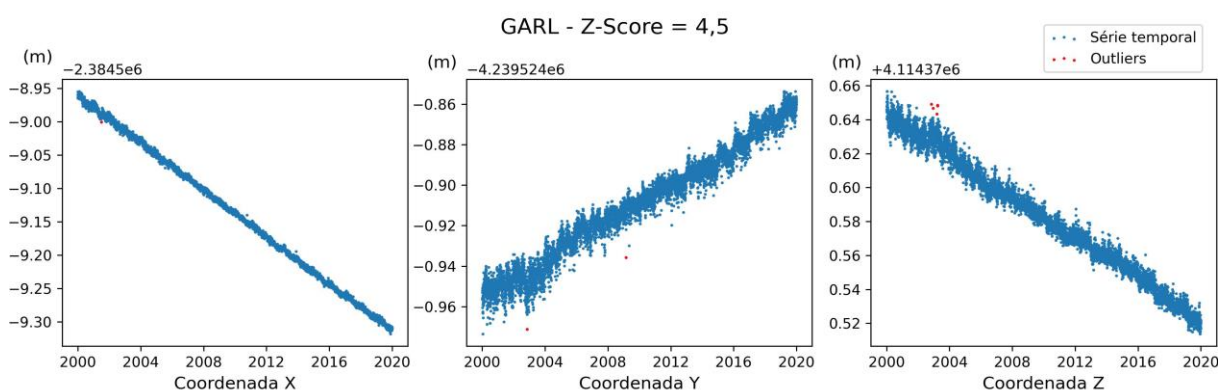
Fonte: elaboração própria.

Das 113 estações, apenas quatro apresentaram redução da discrepância de mais de 1 mm, sendo elas: KOKB, S300, TOST (as mesmas estações que se destacaram no *boxplot* ajustado) e WMAP.

Como esperado, quanto maior o *Z-Score*, mais próxima a discrepância fica da DSRO. Como exceção destaca-se a S300, em que um *Z-Score* de 5 resultou em uma redução mais significativa na discrepância, mas isso se deve ao fato de a série não ser ruidosa e de seus *outliers* serem de grande magnitude e pouco numerosos.

A contaminação média foi de 1,51%, a máxima foi de 6,98%, na KOKB, e em 18 estações a contaminação foi de 0%. Como evidenciado pela contaminação média, o *Z-Score* identificou menos pontos como *outliers* em relação ao *boxplot* ajustado, deixando de identificar diversos pontos que em uma análise visual seriam considerados *outliers*, como exemplificado pela Figura 32.

Figura 32 – *Outliers* identificados nas séries da estação GARL.

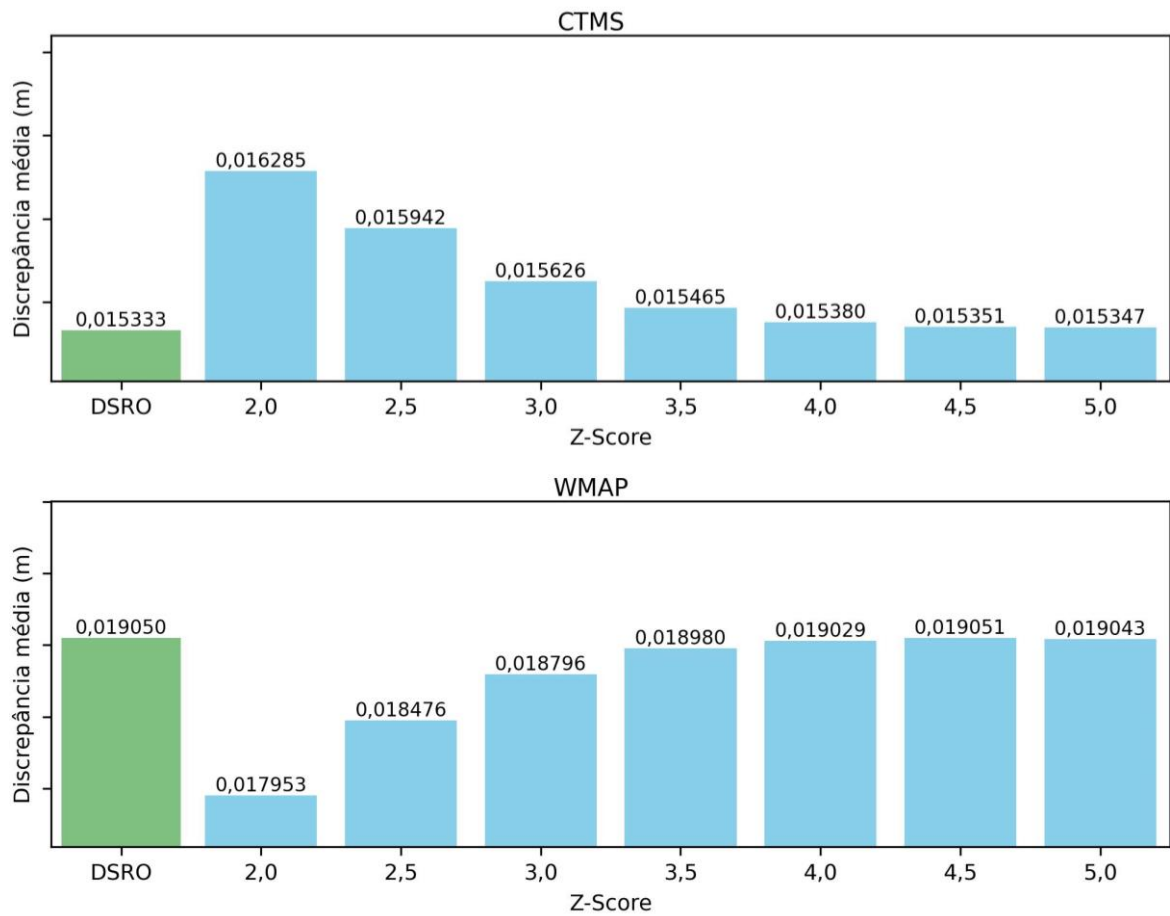


Fonte: elaboração própria.

Salienta-se que, de modo geral, não se observou uma correlação entre a contaminação e uma mudança significativa na discrepância, apesar de isso ter acontecido em estações como a WMAP. A estação S300, com uma contaminação de apenas 0,07%, apresentou a maior redução de discrepância, enquanto estações com alta contaminação como COON (5,45%), OAES (5,54%) e TIVA (5,34%) tiveram redução desprezíveis.

Em relação ao parâmetro *Z-Score* (*threshold*), para a estação CTMS, por exemplo, que possui uma série não ruidosa e com uma tendência constante no período estudado, a remoção dos *outliers* aumentou a discrepância para todos os valores analisados, apesar da diferença para a DSRO ser submilimétrica (Figura 33). O valor do *Z-Score* com melhor resultado foi 5, que implicou em uma contaminação de 0,30%, ou seja, o valor que menos modificou a série original.

Figura 33 - Discrepâncias para os valores de *Z-Score* das estações CTMS e WMAP.



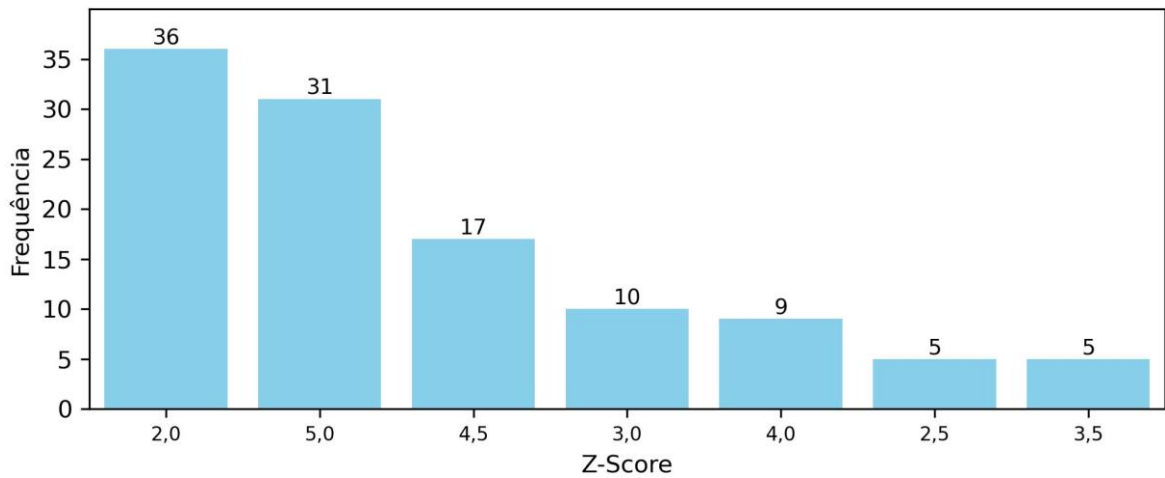
Fonte: elaboração própria.

Já para a WMAP, que como mencionado possui um salto na série, o *Z-Score* de 2, que causa maior interferência na série original, proporcionou a menor discrepância. Ao aumentar o valor do *Z-Score* o valor da discrepância também aumentou, com os valores de *Z-Score* 4,5 e 5 sendo maiores que a DSRO.

Desta forma, o maior impacto na discrepância é causado quando a série possui saltos que divergem da tendência de longa duração ou quando os *outliers* são de magnitude significativas. A remoção de *outliers* de poucos milímetros ou centímetros (para séries com velocidades mais altas) não impactam a velocidade final da estação

O *Z-Score* igual a 2 foi o valor que proporcionou a menor discrepância no maior número de estações analisadas, 36 das 113 (Figura 34). Dessas 36 estações, apenas uma teve discrepância maior que a DSRO, a LMUT, com aumento de 0,000459 m. As outras 35 apresentaram uma redução média de 0,000296 m, valor desprezível.

Figura 34 - Frequência do *Z-Score* que proporcionou a menor discrepância para as 113 estações, após aplicação do *Z-Score*.



Fonte: elaboração própria.

Em 31 estações o *Z-Score* igual a 5 proporcionou a menor discrepância. Contudo, em 14 dessas 31, a discrepância foi maior que a DSRO (comportamento semelhante ao supracitado com a CTSM). Nas outras 17 estações, a redução média foi de 0,000506 m, também não sendo significativa.

Klos *et al.* (2015) utilizaram 3 e 5 vezes o desvio padrão como critério de identificação de *outliers*, Kierulf *et al.* (2021) utilizaram 3σ e Crocetti, Schartner e Soja (2021) 5σ . Como já mencionado, esses valores são equivalentes ao *Z-Score*. O *Z-Score* de 5 condiz com os resultados aqui obtidos, já o valor de 3 não obteve a performance esperada.

De acordo com os resultados, a utilização do *Z-Score* igual a 2 seria indicado para séries ruidosas, com saltos e/ou mudanças de tendências (uma vez que o *Z-Score* mais baixo irá manter apenas os pontos mais próximos da tendência de longa duração e eliminar os saltos e movimentações com tendências de curto período). Já o valor de 5 seria indicado para séries sem estas características, já que irá identificar os *outliers* esporádicos sem maior interferência na tendência.

6.4.4 *Z-Score Modificado*

Após a remoção da tendência das séries e a aplicação do *Z-Score* modificado, observou-se uma redução média na discrepância de 0,000311 m. A maior redução registrada foi de

0,006950 m, verificada na estação S300. Por outro lado, em 30 estações, houve um aumento médio na discrepância de 0,000166 m.

A Tabela 5 mostra a discrepância antes e depois da aplicação do *Z-Score* modificado, qual *threshold* proporcionou a menor discrepância e a contaminação das séries para algumas das estações.

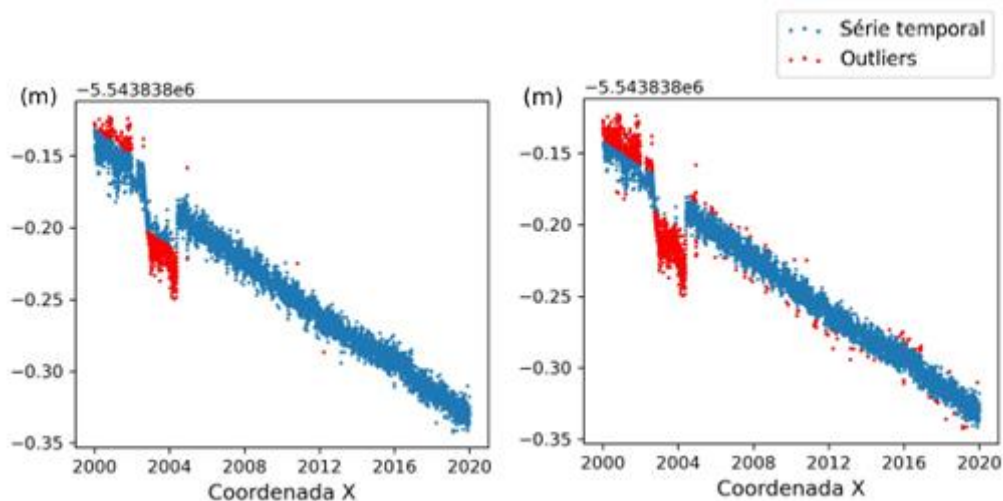
Tabela 5 - Discrepâncias antes e após eliminação de *outliers* com *Z-Score* modificado.

| Estação | Threshold | Discrepância Z-Score Mod. (m) | DSRO (m) | Δ Discrepância (m) | Contaminação (%) |
|---------|-----------|-------------------------------------|-------------|-----------------------|---------------------|
| BARH | 3 | 0,018743 | 0,018779 | 0,000035 | 0,84 |
| BURN | 2 | 0,004338 | 0,004641 | 0,000304 | 6,68 |
| CAND | 3,5 | 0,015416 | 0,015411 | -0,000004 | 0,00 |
| CAST | 2 | 0,003969 | 0,004006 | 0,000038 | 5,60 |
| CJMS | 5 | 0,016344 | 0,015322 | -0,001021 | 1,04 |
| CLAR | 3 | 0,008221 | 0,008226 | 0,000005 | 1,47 |
| CRHS | 5 | 0,009821 | 0,009449 | -0,000371 | 1,39 |
| KNOL | 2 | 0,004989 | 0,005065 | 0,000076 | 4,52 |
| KOKB | 2 | 0,005458 | 0,009953 | 0,004494 | 11,19 |
| LAND | 2 | 0,016858 | 0,016858 | 0,000000 | 0,00 |
| LMUT | 2 | 0,004839 | 0,004498 | -0,000341 | 6,45 |
| MAUI | 2 | 0,003314 | 0,003269 | -0,000044 | 5,96 |
| MIDA | 3 | 0,015175 | 0,015183 | 0,000008 | 0,00 |
| OAES | 2 | 0,015395 | 0,015666 | 0,000271 | 11,40 |
| OPBL | 5 | 0,011413 | 0,011413 | 0,000000 | 0,00 |
| OPCP | 5 | 0,017053 | 0,017045 | -0,000008 | 0,05 |
| ORES | 2,5 | 0,003802 | 0,003802 | 0,000000 | 0,00 |
| OVLS | 4,5 | 0,006798 | 0,006808 | 0,000010 | 0,01 |
| PMTN | 5 | 0,003600 | 0,003848 | 0,000247 | 0,36 |
| POMM | 2 | 0,016350 | 0,016353 | 0,000004 | 0,04 |
| PPBF | 5 | 0,009483 | 0,009479 | -0,000004 | 0,00 |
| PSDM | 2,5 | 0,006131 | 0,006222 | 0,000091 | 2,58 |
| PVHS | 2 | 0,006525 | 0,006636 | 0,000111 | 5,87 |
| S300 | 5 | 0,003006 | 0,009955 | 0,006950 | 0,14 |
| SHIN | 5 | 0,008762 | 0,008762 | 0,000000 | 0,00 |
| TIVA | 2 | 0,003835 | 0,003954 | 0,000120 | 9,75 |
| TOST | 2 | 0,005738 | 0,007375 | 0,001637 | 5,49 |
| WLSN | 5 | 0,008783 | 0,008562 | -0,000221 | 0,91 |
| WMAP | 2 | 0,017642 | 0,019050 | 0,001408 | 6,38 |

Fonte: elaboração própria.

Das 113 estações, apenas quatro apresentaram redução de mais de 1 mm, sendo elas: KOKB, S300, TOST e WMAP, como no *Z-Score*. Destas, apenas a S300 apresentou melhor resultado com o *Z-Score* que com o *Z-Score* modificado. A diferença mais relevante foi na KOKB, que com o *Z-Score* modificado apresentou redução de 0,004494 m, em contraste com os 0,002831 m com o *Z-Score*. Esse resultado é explicado pelo *Z-Score* modificado ter sido mais restritivo, identificando um número maior de pontos como *outliers*, como mostrado na Figura 35.

Figura 35 – Série de coordenada X da estação KOKB, com *outliers* identificados com *Z-Score* (esquerda) e *Z-Score* modificado (direita).

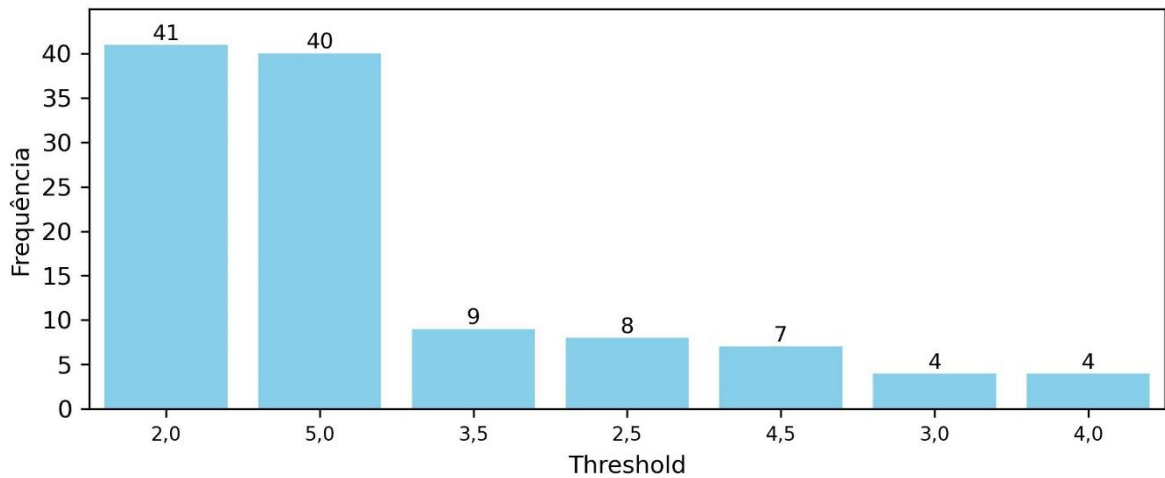


Fonte: elaboração própria.

Esse comportamento mais restritivo também foi verificado nas demais 112 estações, o que foi refletido na contaminação média, 2,59% em contraste com os 1,51% do *Z-Score*. A contaminação máxima foi de 11,40% na OAES e a mínima de 0%, em 17 estações.

Iglewicz e Hoaglin (1993) sugerem um *threshold* de 3,5 para a aplicação do *Z-Score* modificado para dados em geral. No entanto, os resultados apontam que os valores de 2 (melhor em 41 estações) e 5 (melhor em 40 estações) são os mais indicados para séries temporais posicionais, conforme ilustrado na Figura 36.

Figura 36 - Frequência de *threshold* que proporcionou a menor discrepância após aplicação do *Z-Score* modificado.



Fonte: elaboração própria.

Destaca-se que para a maioria das estações a diferença entre o *Z-Score* e o *Z-Score* modificado não foi relevante, apesar de o *Z-Score* modificado geralmente apresentar melhores resultados.

6.4.5 BCP-HI

Uma vez aplicado o BCP-HI, notou-se uma redução média na discrepância de 0,000229 m. A maior redução registrada foi de 0,007718 m, verificada na estação S300. Por outro lado, em 4 estações, houve um aumento médio na discrepância de 0,000556 m.

A Tabela 6 mostra a discrepância antes e depois da aplicação do BCP-HI, qual cp (*change points*) proporcionou a menor discrepância e a contaminação das séries para algumas das estações.

Tabela 6 - Discrepâncias antes e após eliminação de *outliers* com BCP-HI.

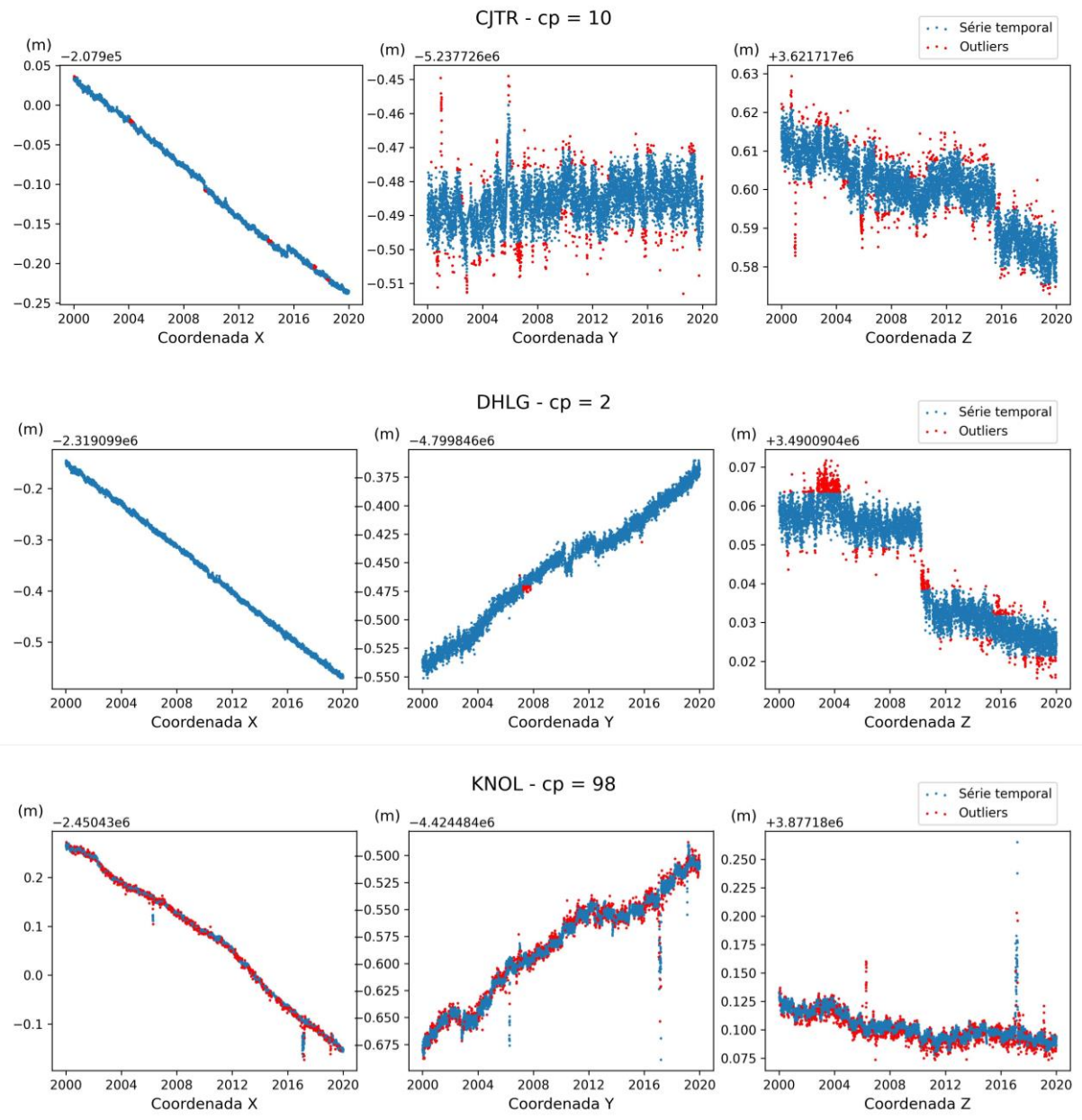
| Estação | <i>cp</i> | Discrepância BCP-HI (m) | DSRO (m) | Δ Discrepância (m) | Contaminação (%) |
|---------|-----------|----------------------------|-------------|------------------------------|---------------------|
| BARH | 100 | 0,018746 | 0,018779 | 0,000033 | 7,99 |
| BURN | 22 | 0,004463 | 0,004641 | 0,000178 | 5,41 |
| CAND | 40 | 0,015297 | 0,015411 | 0,000114 | 5,19 |
| CAST | 16 | 0,003787 | 0,004006 | 0,000220 | 5,69 |
| CJMS | 42 | 0,015277 | 0,015322 | 0,000045 | 5,64 |
| CLAR | 12 | 0,007359 | 0,008226 | 0,000867 | 1,52 |
| CRHS | 14 | 0,009303 | 0,009449 | 0,000146 | 2,03 |
| KNOL | 98 | 0,004897 | 0,005065 | 0,000168 | 8,14 |
| KOKB | 96 | 0,009824 | 0,009953 | 0,000129 | 5,13 |
| LAND | 14 | 0,016732 | 0,016858 | 0,000125 | 1,58 |
| LMUT | 8 | 0,004875 | 0,004498 | -0,000376 | 2,82 |
| MAUI | 6 | 0,003204 | 0,003269 | 0,000066 | 0,14 |
| MIDA | 56 | 0,015082 | 0,015183 | 0,000101 | 5,74 |
| OAES | 4 | 0,014678 | 0,015666 | 0,000988 | 2,40 |
| OPBL | 28 | 0,011348 | 0,011413 | 0,000065 | 4,89 |
| OPCP | 14 | 0,016906 | 0,017045 | 0,000139 | 4,44 |
| ORES | 4 | 0,002971 | 0,003802 | 0,000831 | 1,62 |
| OVLS | 96 | 0,006756 | 0,006808 | 0,000053 | 6,16 |
| PMTN | 16 | 0,003333 | 0,003848 | 0,000515 | 4,90 |
| POMM | 14 | 0,016234 | 0,016353 | 0,000119 | 2,37 |
| PPBF | 20 | 0,009327 | 0,009479 | 0,000151 | 2,04 |
| PSDM | 12 | 0,006192 | 0,006222 | 0,000030 | 1,01 |
| PVHS | 20 | 0,006410 | 0,006636 | 0,000226 | 2,35 |
| S300 | 44 | 0,002237 | 0,009955 | 0,007718 | 4,99 |
| SHIN | 52 | 0,008656 | 0,008762 | 0,000106 | 6,18 |
| TIVA | 16 | 0,003805 | 0,003954 | 0,000149 | 3,98 |
| TOST | 12 | 0,007360 | 0,007375 | 0,000015 | 0,77 |
| WLSN | 28 | 0,008397 | 0,008562 | 0,000166 | 5,17 |
| WMAP | 100 | 0,018985 | 0,019050 | 0,000065 | 5,86 |

Fonte: elaboração própria.

A S300 foi a única estação que apresentou uma redução de mais de 1 mm na discrepância (especificamente de 0,007718 m) após a aplicação do BCP-HI.

Os resultados mostraram que o método ignorou os saltos nas séries e não os tratou como *outliers*, como exemplificado pelas estações CJTR e DHLG (Figura 37). Por outro lado, o método falhou em identificar alguns *outliers* de grande magnitude situados dentro de agrupamentos de pontos deslocados da tendência de longa duração, como na série da estação KNOL.

Figura 37 – *Outliers* detectados nas séries das estações CJTR, DHLG e KNOL com CBP-HI.

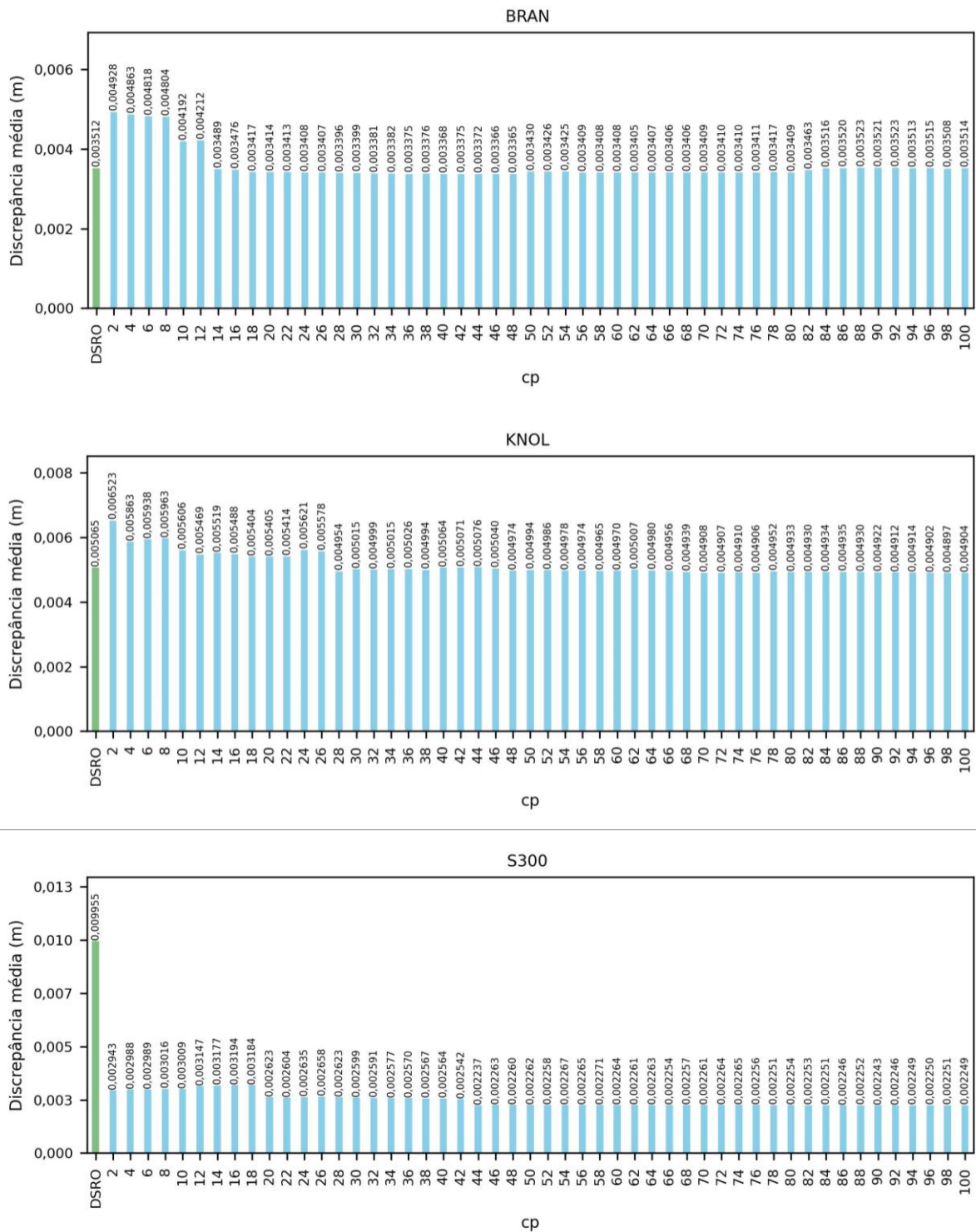


Fonte: elaboração própria.

Considerando as 113 estações analisadas, a contaminação média foi de 3,81%. O valor máximo foi de 8,14%, registrado na estação KNOL, enquanto o valor mínimo (0%) foi observado na estação LORS.

A variação dos valores do cp na maioria das estações não resultou em uma variação relevante na discrepância, mas nas estações em que isso ocorreu, como na BRAN, KNOL e S300 (Figura 38), um cp de 44 já foi o suficiente para obter a discrepância mínima.

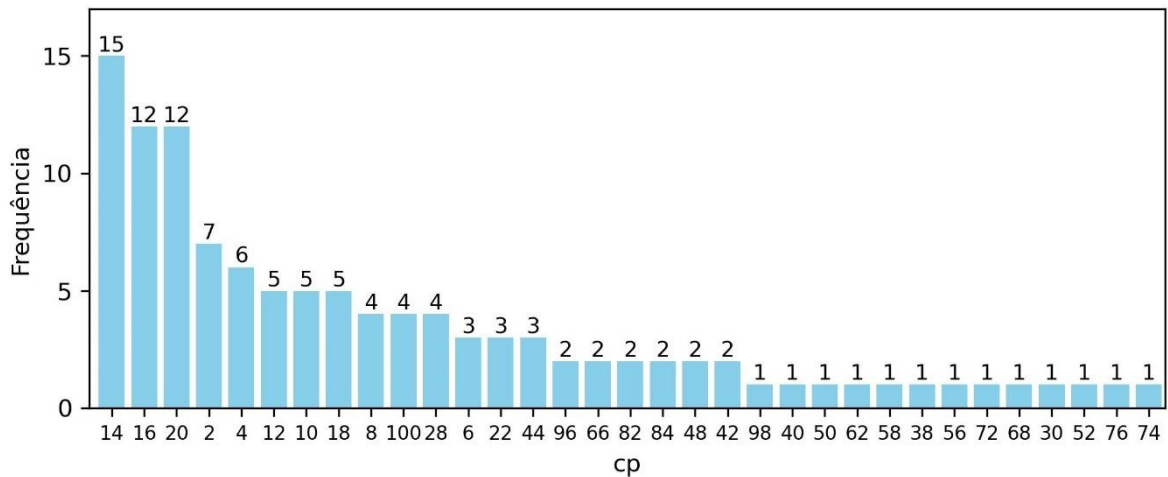
Figura 38 - Valores de cp e suas discrepâncias correspondentes.



Fonte: elaboração própria.

Entre os 9 valores de cp que proporcionaram a menor discrepância no maior número de estações, todos são menores ou igual a 20, totalizando 71 estações (62,8%), elevando o cp para 44, mais 16 estações são incluídas, como mostra a Figura 39.

Figura 39 - Frequência de cp que proporcionou a menor discrepância após aplicação BCP-HI.



Fonte: elaboração própria.

Pehlivan (2024) analisando 3 séries de dados sintéticos verificou o valor ótimo de cp para cada uma delas como sendo 20, 42 e 120. Já para séries de dados reais, o cp ótimo foi de 52. Deste modo, um valor de cp entre 44 e 52 pode ser considerado de forma geral como indicado para aplicação em séries temporais posicionais reais.

6.4.6 KNN

Após a remoção da tendência das séries e a aplicação do KNN, observou-se uma redução média na discrepância de 0,000372 m. A maior redução registrada foi de 0,006867 m, verificada na estação S300. No entanto, em 44 estações, houve um aumento médio na discrepância de 0,000119 m.

A Tabela 7 mostra a discrepância antes e depois da aplicação do KNN e quais parâmetros proporcionaram a menor discrepância para algumas das estações.

Tabela 7 - Discrepâncias antes e após eliminação de *outliers* com KNN.

| Estação | Nº vizinhos | Contaminação (%) | Discrepância KNN (m) | DSRO (m) | Δ Discrepância (m) |
|---------|-------------|------------------|----------------------|----------|---------------------------|
| BARH | 4 | 3 | 0,018706 | 0,018779 | 0,000073 |
| BURN | 33 | 7 | 0,004205 | 0,004641 | 0,000436 |
| CAND | 19 | 10 | 0,015016 | 0,015411 | 0,000395 |
| CAST | 87 | 9 | 0,003934 | 0,004006 | 0,000072 |
| CJMS | 3 | 1 | 0,015712 | 0,015322 | -0,000390 |
| CLAR | 4 | 2 | 0,007477 | 0,008226 | 0,000749 |
| CRHS | 14 | 3 | 0,009369 | 0,009449 | 0,000081 |
| KNOL | 2 | 2 | 0,004592 | 0,005065 | 0,000473 |
| KOKB | 22 | 1 | 0,010064 | 0,009953 | -0,000112 |
| LAND | 59 | 10 | 0,015575 | 0,016858 | 0,001283 |
| LMUT | 7 | 10 | 0,004750 | 0,004498 | -0,000252 |
| MAUI | 9 | 10 | 0,003273 | 0,003269 | -0,000004 |
| MIDA | 90 | 10 | 0,013963 | 0,015183 | 0,001220 |
| OAES | 80 | 10 | 0,014828 | 0,015666 | 0,000838 |
| OPBL | 3 | 1 | 0,011504 | 0,011413 | -0,000092 |
| OPCP | 2 | 10 | 0,016982 | 0,017045 | 0,000062 |
| ORES | 3 | 9 | 0,003749 | 0,003802 | 0,000053 |
| OVLS | 40 | 3 | 0,006811 | 0,006808 | -0,000003 |
| PMTN | 3 | 1 | 0,003575 | 0,003848 | 0,000273 |
| POMM | 74 | 10 | 0,015739 | 0,016353 | 0,000615 |
| PPBF | 33 | 1 | 0,009541 | 0,009479 | -0,000062 |
| PSDM | 72 | 10 | 0,006137 | 0,006222 | 0,000085 |
| PVHS | 61 | 9 | 0,006497 | 0,006636 | 0,000139 |
| S300 | 2 | 1 | 0,003088 | 0,009955 | 0,006867 |
| SHIN | 2 | 1 | 0,008797 | 0,008762 | -0,000035 |
| TIVA | 45 | 10 | 0,003747 | 0,003954 | 0,000207 |
| TOST | 32 | 3 | 0,005504 | 0,007375 | 0,001871 |
| WLSN | 2 | 1 | 0,008762 | 0,008562 | -0,000200 |
| WMAP | 74 | 10 | 0,017427 | 0,019050 | 0,001623 |

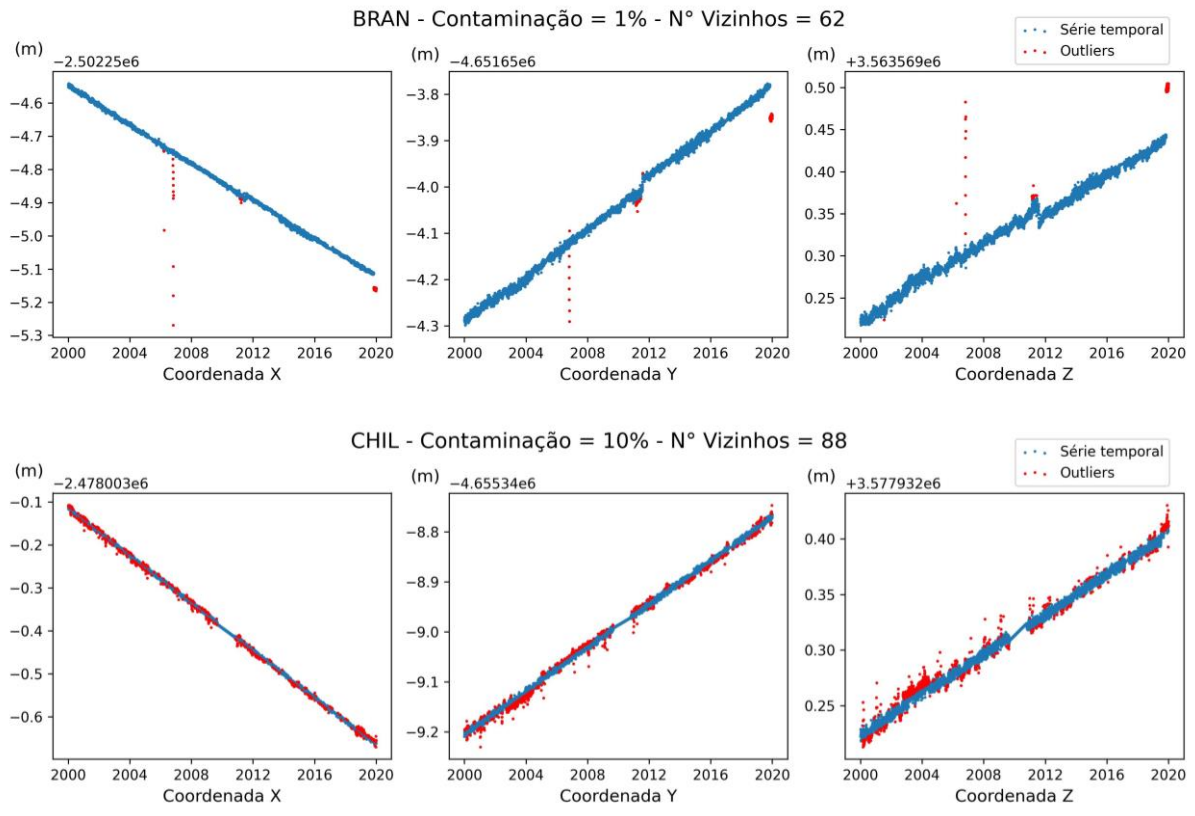
Fonte: elaboração própria.

Das 113 estações, apenas 5 apresentaram redução de mais de 1 mm, sendo elas: S300, TOST e WMAP, como no *Z-Score* e *Z-Score* modificado, além da LAND e MIDA. Como destacado no item 6.2 (página 49), as estações LAND e MIDA possuem um salto na série devido a um terremoto em 2004.

O KNN foi capaz de detectar os *outliers* com uma exatidão superior aos métodos estatísticos, também sendo muito consistente ao detectar agrupamentos de *outliers*, característica que o BCP-HI apresentou dificuldades em lidar.

A Figura 40 exemplifica a capacidade do método não só de identificar os *outliers*, mas também de empregar a contaminação ao longo da série, como o 1% na estação BRAN e 10% na CHIL.

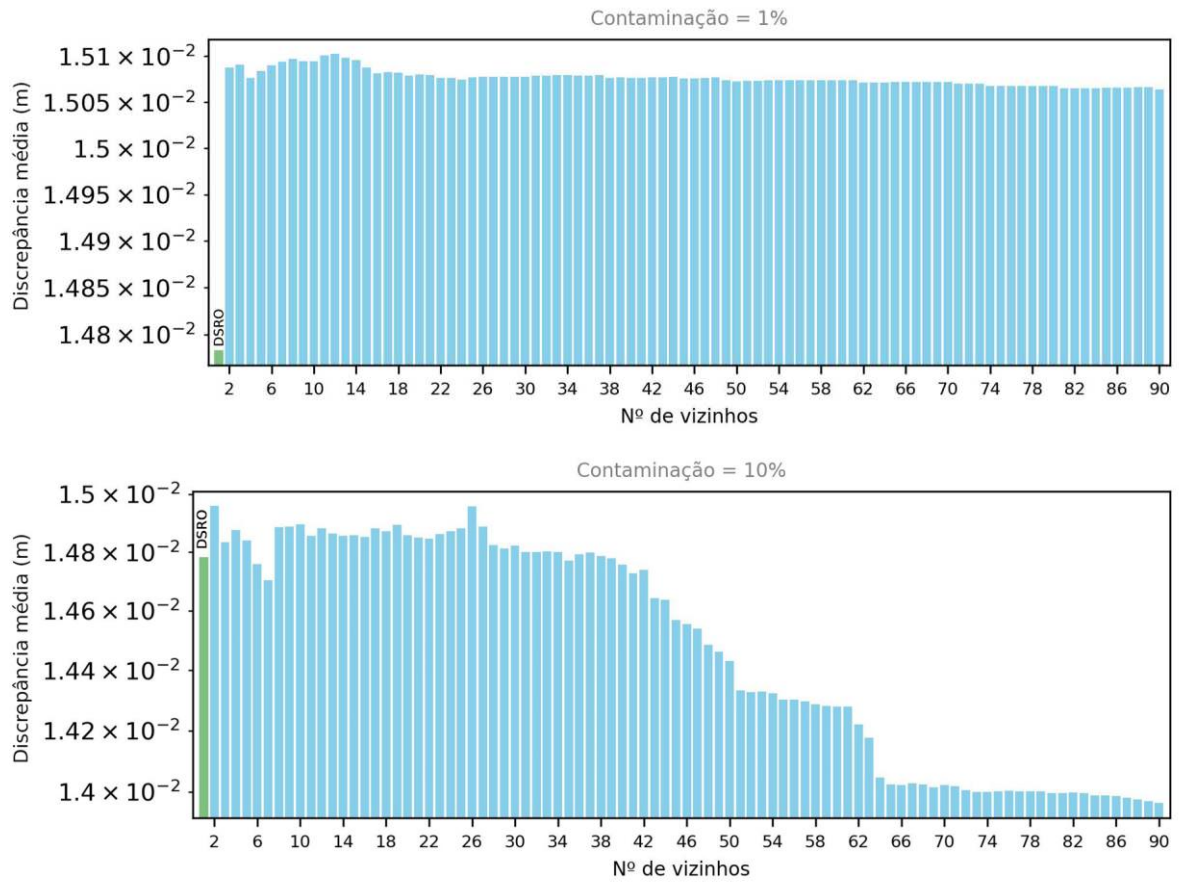
Figura 40 - *Outliers* identificados com KNN.



Fonte: elaboração própria.

Na maioria das séries, a discrepância foi pouco afetada tanto pela variação na contaminação quanto pelo número de vizinhos. As exceções foram as estações com saltos ou mudanças locais na tendência de longa período, como na LAND, MIDA e WMAP. Na MIDA, por exemplo, a pior discrepância com 1% de contaminação foi 0,015102 m (com 12 vizinhos), e a melhor com 10% de contaminação foi 0,013963 m (com 90 vizinhos), uma diferença de 0,001139 m, como ilustrado na Figura 41.

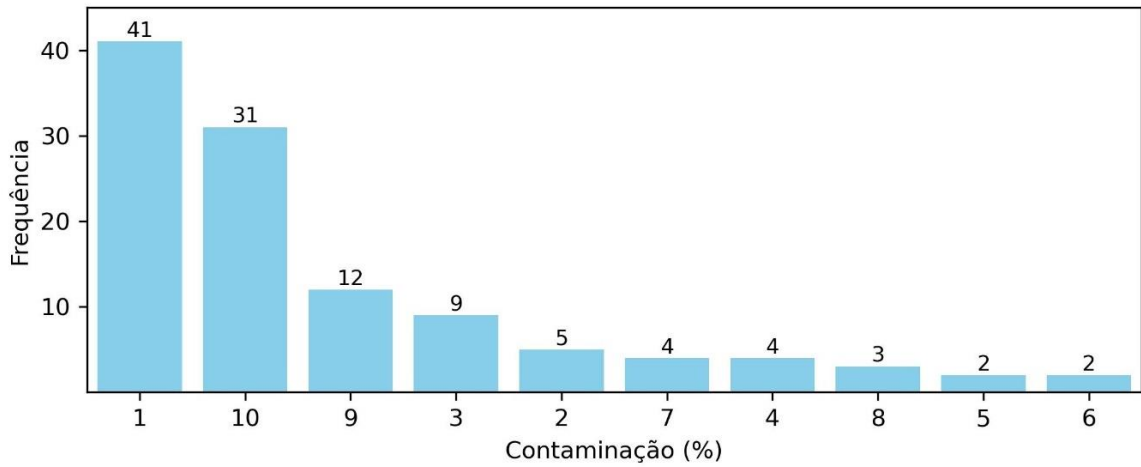
Figura 41 - Discrepâncias com a variação da contaminação e do número de vizinhos para a estação MIDA.



Fonte: elaboração própria.

Analisando quais contaminações proporcionaram a menor discrepância em cada estação, observa-se a maior frequência dos valores de 1 e de 10% (Figura 42), melhores em 41 e 31 estações, respectivamente. Somados, essas duas contaminações compreendem 63,7% das estações. Como já mencionado, as contaminações de 10% estão mais associadas a séries com saltos.

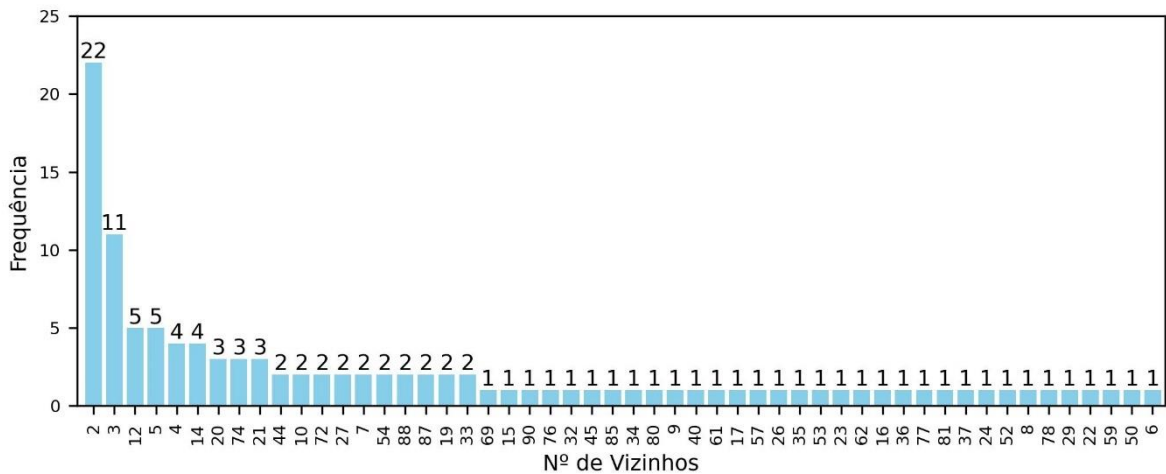
Figura 42 - Frequência da contaminação que proporcionou a menor discrepância após aplicação do KNN.



Fonte: elaboração própria.

Quanto ao número de vizinhos, apesar da maior frequência ter sido de 2 e 3 (Figura 43), quando a mudança do valor deste parâmetro efetivamente afetou a discrepância, valores mais altos proporcionaram menores discrepâncias, como 59 na LAND, 90 na MIDA e 74 na WMAP.

Figura 43 - Frequência dos números de vizinhos que proporcionaram as menores discrepâncias após aplicação do KNN.



Fonte: elaboração própria.

Desse modo, os valores de contaminação de 1% com 2 ou 3 vizinhos se mostraram mais indicados para séries sem problemas. Para séries com problemas, a contaminação de 10% com um número de vizinhos em torno de 90 seria mais apropriado.

6.4.7 Isolation Forest

Após a remoção da tendência das séries e a aplicação do IF, observou-se uma redução média na discrepância de 0,000437 m. A maior redução registrada foi de 0,006897 m, verificada na estação S300. No entanto, em 56 estações, houve um aumento médio na discrepância de 0,000128 m.

A Tabela 8 mostra a discrepância antes e depois da aplicação do IF e qual parâmetro proporcionou a menor discrepância para algumas das estações.

Tabela 8 - Discrepâncias antes e após eliminação de *outliers* com IF.

| Estação | Contaminação (%) | Discrepância IF (m) | DSRO (m) | Δ Discrepância (m) |
|---------|------------------|---------------------|----------|---------------------------|
| BARH | 9 | 0,018757 | 0,018779 | 0,000022 |
| BURN | 8 | 0,004209 | 0,004641 | 0,000432 |
| CAND | 10 | 0,014530 | 0,015411 | 0,000881 |
| CAST | 8 | 0,003938 | 0,004006 | 0,000068 |
| CJMS | 1 | 0,016084 | 0,015322 | -0,000762 |
| CLAR | 2 | 0,008241 | 0,008226 | -0,000015 |
| CRHS | 1 | 0,009518 | 0,009449 | -0,000068 |
| KNOL | 1 | 0,004707 | 0,005065 | 0,000358 |
| KOKB | 10 | 0,009236 | 0,009953 | 0,000717 |
| LAND | 10 | 0,014712 | 0,016858 | 0,002146 |
| LMUT | 8 | 0,004818 | 0,004498 | -0,000320 |
| MAUI | 9 | 0,003282 | 0,003269 | -0,000012 |
| MIDA | 10 | 0,013066 | 0,015183 | 0,002117 |
| OAES | 10 | 0,015217 | 0,015666 | 0,000448 |
| OPBL | 1 | 0,011508 | 0,011413 | -0,000095 |
| OPCP | 1 | 0,017192 | 0,017045 | -0,000148 |
| ORES | 10 | 0,003707 | 0,003802 | 0,000095 |
| OVLS | 1 | 0,006831 | 0,006808 | -0,000022 |
| PMTN | 1 | 0,003763 | 0,003848 | 0,000085 |
| POMM | 10 | 0,015263 | 0,016353 | 0,001090 |
| PPBF | 1 | 0,009542 | 0,009479 | -0,000064 |
| PSDM | 10 | 0,006171 | 0,006222 | 0,000051 |
| PVHS | 10 | 0,006564 | 0,006636 | 0,000072 |
| S300 | 1 | 0,003058 | 0,009955 | 0,006897 |
| SHIN | 10 | 0,008835 | 0,008762 | -0,000073 |
| TIVA | 9 | 0,003795 | 0,003954 | 0,000159 |
| TOST | 4 | 0,005575 | 0,007375 | 0,001800 |
| WLSN | 1 | 0,008838 | 0,008562 | -0,000275 |
| WMAP | 10 | 0,017509 | 0,019050 | 0,001541 |

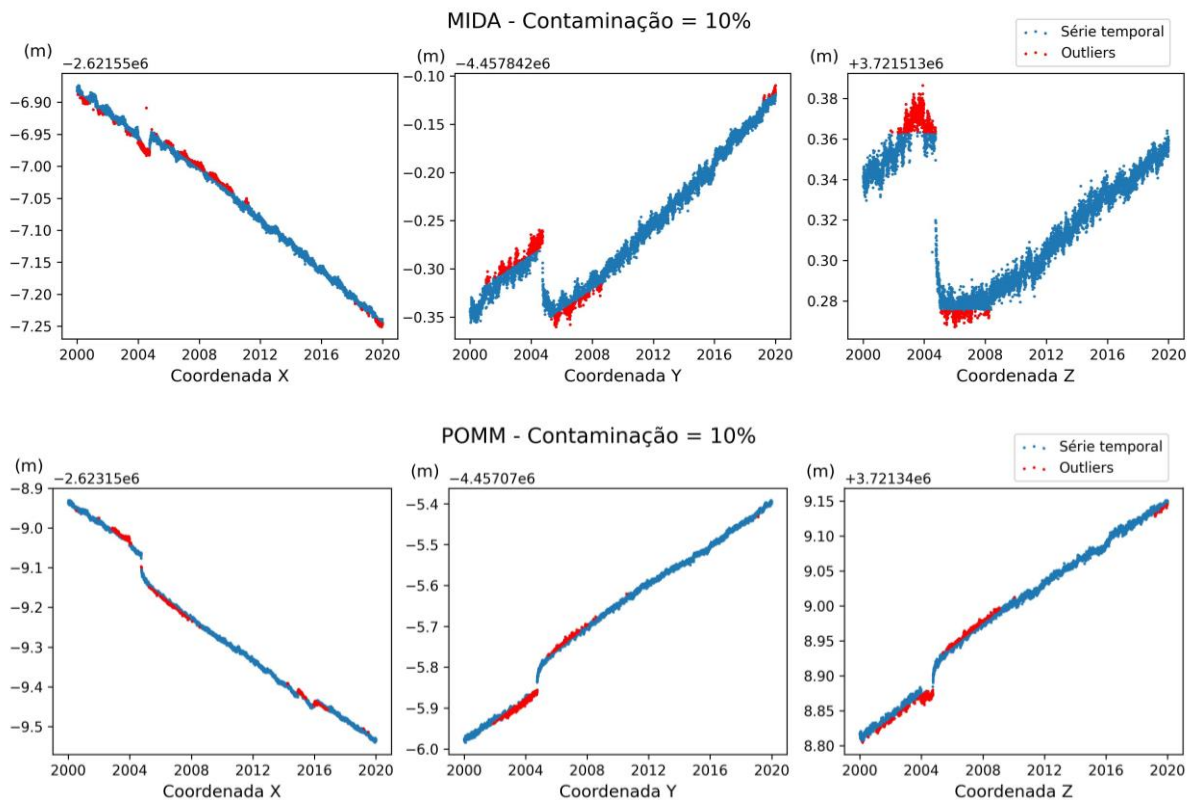
Fonte: elaboração própria.

Das 113 estações, apenas 6 apresentaram redução de discrepância de mais de 1 mm, sendo elas:

- TOST e S300: como no *boxplot* ajustado, *Z-Score*, *Z-Score* modificado e KNN;
- WMAP: como no *Z-Score*, *Z-Score* modificado e KNN;
- LAND e MIDA: como no KNN; e
- POMM: como no *boxplot* ajustado.

Como pode ser observado na Figura 44, em séries com saltos, a remoção dos *outliers* identificados perto das discontinuidades contribuíram para uma maior linearização da série em relação ao ocorrido em séries sem saltos e, conseqüentemente, apresentaram um impacto positivo na discrepância. Entretanto, destaca-se que muitos desses pontos não são verdadeiramente *outliers*, mas pontos que sofreram deslocamentos devido a atividades sísmicas.

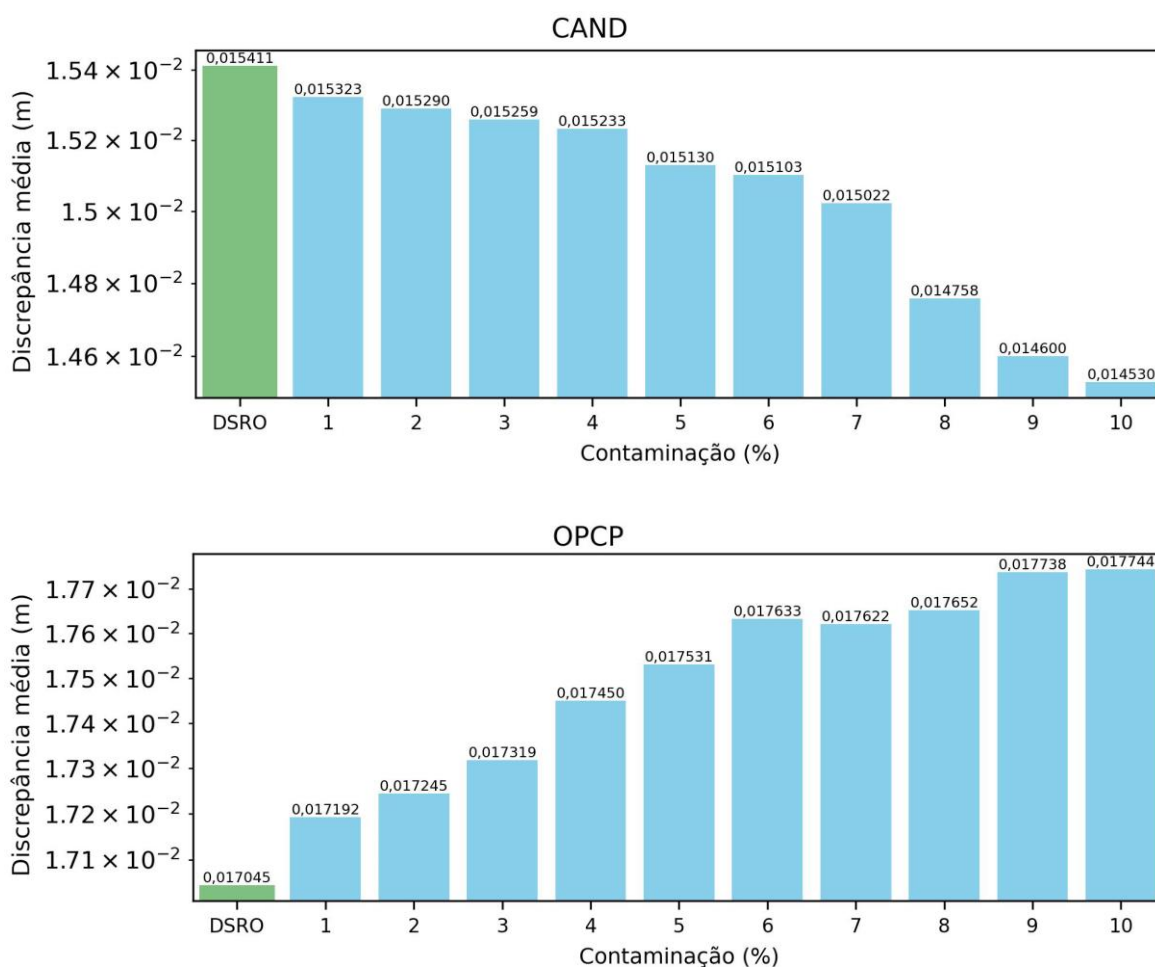
Figura 44 - *Outliers* identificados com IF nas estações MIDA e POMM.



Fonte: elaboração própria.

A variação da contaminação apresentou um comportamento majoritariamente crescente ou decrescente nas discrepâncias. Em outras palavras, para algumas estações quanto maior foi a contaminação, maior foi a discrepância. E para outras, quanto maior a contaminação, menor a discrepância, como ilustrado pelas estações CAND e OPCP (Figura 45). O padrão que se pôde observar foi o mesmo de outros métodos, quando a contaminação mais alta melhorou a discrepância, notou-se a presença de saltos nas séries.

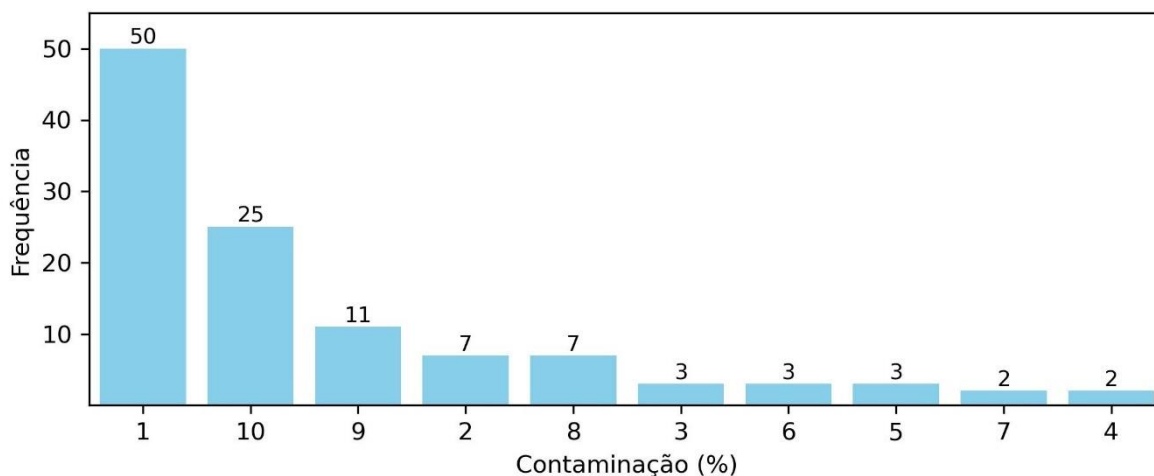
Figura 45 - Relação da discrepância com a variação da contaminação com o IF nas estações CAND e OPCP.



Fonte: elaboração própria.

Como ocorrido no KNN, com o IF as contaminações que proporcionaram as menores discrepâncias foram de 1 e de 10%, em 66,4% das estações, como mostra a Figura 46. Entretanto, no IF houve uma predominância maior da contaminação de 1%.

Figura 46 - Frequência da contaminação que proporcionou a menor discrepância após aplicação do IF.



Fonte: elaboração própria.

Assim como visto nos outros métodos, contaminação de 10% ficou mais associada às séries com saltos e mudanças de tendências, como a CAND, LAND, MIDA etc. E as contaminações menores ficaram mais relacionadas às séries sem problemas ou com pequenos grupos de *outliers*, como ELSC, GMRC, LMUT, entre outras.

Nguyen e Tran (2023) utilizaram séries sintéticas de 3 anos (1095 coordenadas contínuas) para avaliar a detecção de *outliers* com IF. O conjunto de dados foi criado gerando uma sequência de coordenadas base em uma equação que inclui deslocamento linear e variação sazonal. Foram adicionados erros aleatórios a todas as coordenadas, com desvio padrão de ± 1 mm e obedecendo uma distribuição normal. Além disso, foram introduzidos erros maiores, variando de ± 10 mm a ± 200 mm, em 164 coordenadas (15% da sequência) para gerar *outliers*. Segundo os resultados encontrados, o IF foi eficaz e confiável, com precisão de 93% na detecção de *outliers*, superando métodos como *One-Class Support Vector Machines* (O-C SVM) e *Local Outlier Factor* (LOF).

6.4.8 DBSCAN - Density-Based Spatial Clustering of Applications with Noise

Após a remoção da tendência das séries e a aplicação do KNN, observou-se uma redução média na discrepância de 0,000835 m. A maior redução registrada foi de 0,012505 m, verificada na estação S300. Por outro lado, em 14 estações, houve um aumento médio na discrepância de 0,000101 m.

A Tabela 9 apresenta a discrepância antes e depois da aplicação do DBSCAN e quais parâmetros (número mínimo de pontos e épsilon) proporcionaram a menor discrepância para algumas das estações.

Tabela 9 - Discrepâncias antes e após eliminação de *outliers* com DBSCAN.

| Estação | Nº mínimo de pontos da amostra | Épsilon (m) | Discrepância DBSCAN (m) | DSRO (m) | Δ Discrepância (m) | Contaminação (%) |
|---------|--|----------------|-------------------------------|-------------|---------------------------------|---------------------|
| BARH | 8 | 0,0001 | 0,018748 | 0,018779 | 0,000031 | 1,15 |
| BURN | 87 | 0,0001 | 0,004082 | 0,004641 | 0,000560 | 28,93 |
| CAND | 37 | 0,0001 | 0,004003 | 0,015411 | 0,011408 | 61,95 |
| CAST | 63 | 0,0001 | 0,003899 | 0,004006 | 0,000107 | 15,7 |
| CJMS | 2 | 0,0007 | 0,015339 | 0,015322 | -0,000017 | 0,04 |
| CLAR | 16 | 0,0007 | 0,007210 | 0,008226 | 0,001016 | 0,84 |
| CRHS | 4 | 0,0005 | 0,009372 | 0,009449 | 0,000077 | 0,25 |
| KNOL | 69 | 0,0001 | 0,003706 | 0,005065 | 0,001359 | 70,34 |
| KOKB | 77 | 0,0005 | 0,004388 | 0,009953 | 0,005565 | 10,6 |
| LAND | 38 | 0,0001 | 0,004352 | 0,016858 | 0,012505 | 72,62 |
| LMUT | 90 | 0,0001 | 0,004742 | 0,004498 | -0,000243 | 24,81 |
| MAUI | 90 | 0,0001 | 0,003274 | 0,003269 | -0,000005 | 22,13 |
| MIDA | 31 | 0,0001 | 0,004783 | 0,015183 | 0,010400 | 66,78 |
| OAES | 50 | 0,0001 | 0,014344 | 0,015666 | 0,001322 | 31,15 |
| OPBL | 3 | 0,0009 | 0,011409 | 0,011413 | 0,000004 | 0,03 |
| OPCP | 83 | 0,0001 | 0,015968 | 0,017045 | 0,001076 | 37,13 |
| ORES | 52 | 0,0001 | 0,002436 | 0,003802 | 0,001365 | 76,46 |
| OVLS | 2 | 0,0011 | 0,006798 | 0,006808 | 0,000010 | 0,01 |
| PMTN | 84 | 0,0001 | 0,003366 | 0,003848 | 0,000482 | 56,8 |
| POMM | 39 | 0,0001 | 0,008598 | 0,016353 | 0,007755 | 61,55 |
| PPBF | 3 | 0,0007 | 0,009474 | 0,009479 | 0,000005 | 0,04 |
| PSDM | 57 | 0,0001 | 0,006128 | 0,006222 | 0,000094 | 14,52 |
| PVHS | 29 | 0,0001 | 0,006427 | 0,006636 | 0,000209 | 7,83 |
| S300 | 4 | 0,0005 | 0,002999 | 0,009955 | 0,006956 | 0,24 |
| SHIN | 2 | 0,0013 | 0,008763 | 0,008762 | -0,000001 | 0,00 |
| TIVA | 90 | 0,0001 | 0,003726 | 0,003954 | 0,000229 | 22,56 |
| TOST | 12 | 0,0001 | 0,005519 | 0,007375 | 0,001855 | 2,93 |
| WLSN | 2 | 0,0015 | 0,008593 | 0,008562 | -0,000031 | 0,09 |
| WMAP | 63 | 0,0001 | 0,013423 | 0,019050 | 0,005627 | 47,82 |

Fonte: elaboração própria.

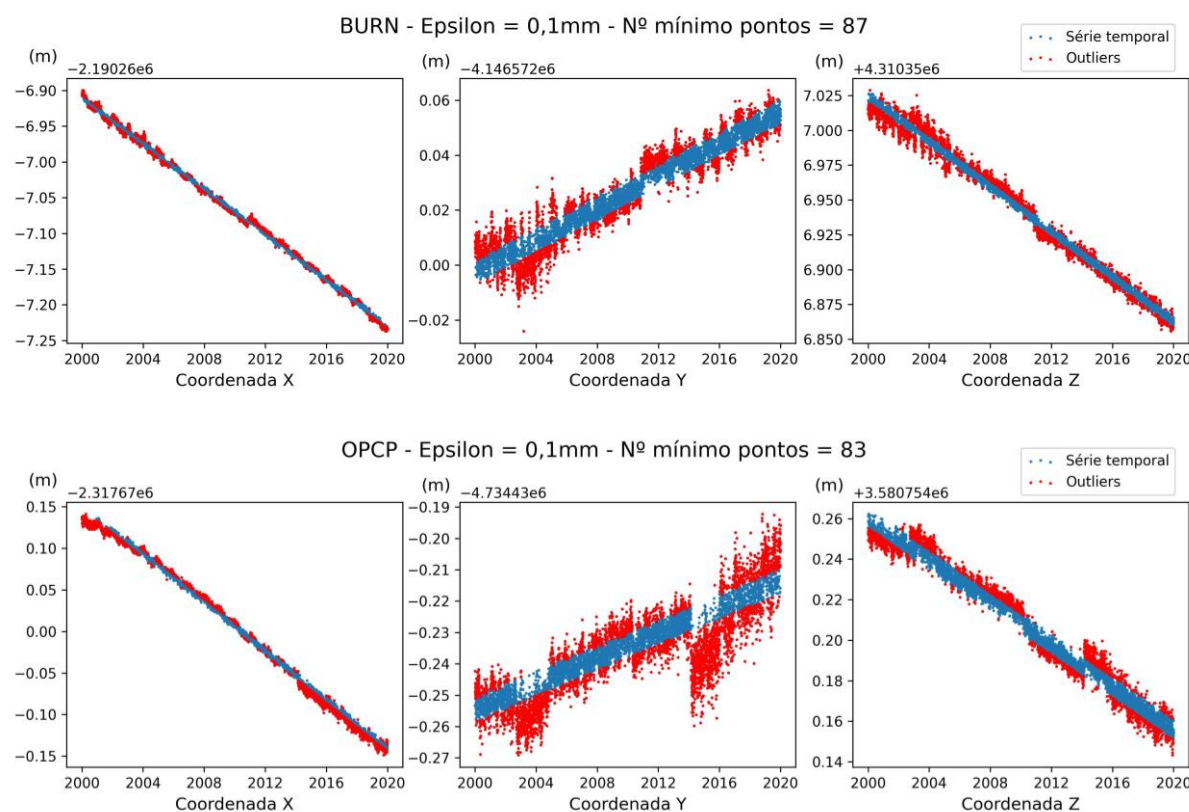
Das 113 estações, 13 apresentaram redução de discrepância de mais de 1 mm, sendo elas:

- TOST e S300: como no *boxplot* ajustado, *Z-Score*, *Z-Score* modificado, KNN e IF;
- WMAP: como no *Z-Score*, *Z-Score* modificado, KNN e IF;
- LAND e MIDA: como no KNN e IF;
- POMM: como no *boxplot* ajustado e IF;
- KOKB: como no *boxplot* ajustado e *Z-Score* modificado.; e,
- CAND, CLAR, OAES, KNOL, OPCP e ORES: somente no DBSCAN.

Dentre as estações CAND, CLAR, OAES, KNOL, OPCP e ORES, a CAND possui um salto devido a um terremoto, como já mencionado. A CLAR passou por uma troca de antena e receptor em novembro de 2019. A OAES passou por troca de antena em outubro de 2000 e setembro de 2007, além de um terremoto em abril de 2010 (7,2 na escala Richter a 218,4 km da estação) e dois em julho de 2019 (magnitude 6,4 a 217,7 km e 7,1 a 228,6 km). A KNOL passou por troca de antena em outubro de 2002 e troca de receptor em junho de 2011, além de 5 terremotos em dezembro de 2016 e julho de 2019 (com magnitudes entre 5,5 e 7,1, com epicentro distante entre 80 e 254 km da estação). A OPCP e a ORES apesar de possuírem trocas de antena e receptor e registro de terremotos no período estudado, nenhum desses eventos coincide com os saltos presentes nas séries.

Além de melhorar a discrepância em mais de 0,001 m em todas as estações que os outros métodos melhoraram, o DBSCAN expandiu tais resultados para mais 6 estações, sendo o método que a apresentou melhores resultados em termos de discrepância entre todos os analisados. Uma característica que pode explicar esta melhora é o aspecto mais restritivo da identificação de *outliers*, em quem os “limites superiores e inferiores” em relação à linha de tendência são menores em relação aos outros métodos. Tal aspecto pode ser observado na Figura 47.

Figura 47 - *Outliers* identificados com DBSCAN nas séries das estações BURN e OPCP.



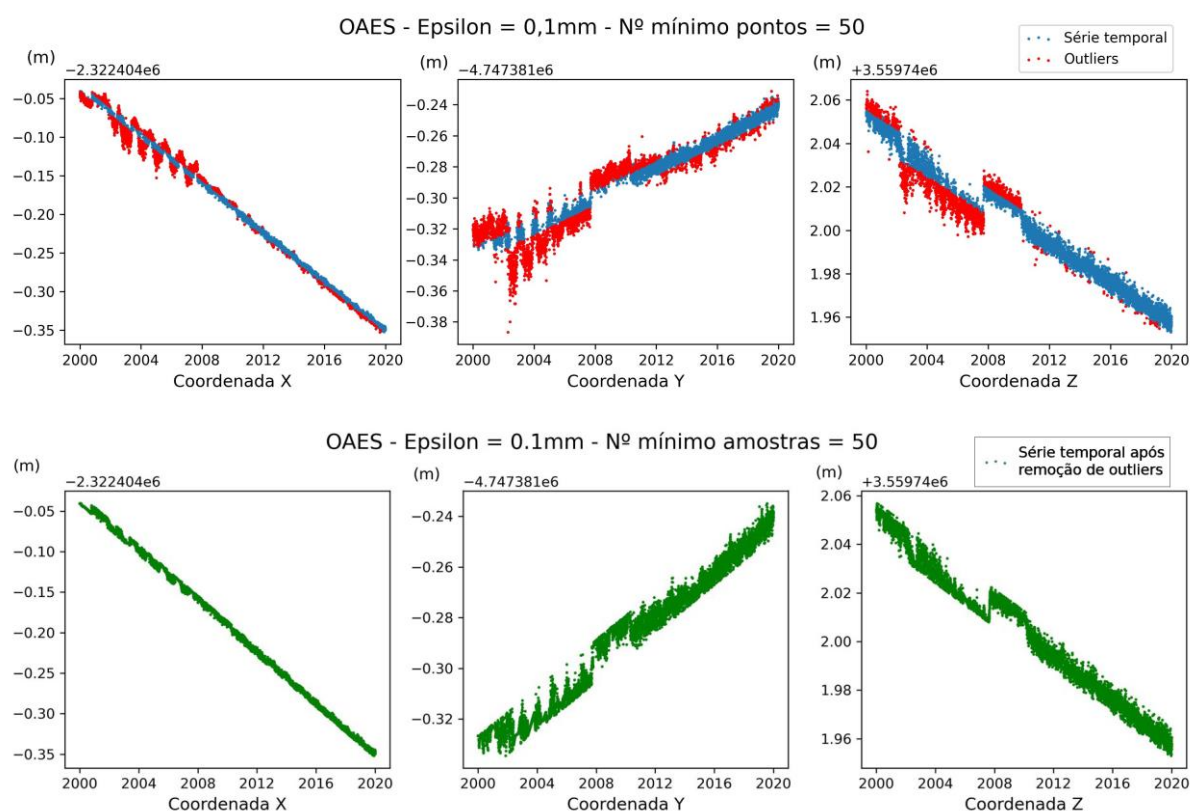
Fonte: elaboração própria.

Todavia, deve-se atentar à dificuldade de calibração do método para aplicação em várias séries ao mesmo tempo, o que ocasionou uma grande divergência nos valores de contaminação entre as séries, ainda que com os mesmos parâmetros. Segundo Peschon (2020), a atribuição de valores aos parâmetros épsilon e número mínimo de pontos na amostra é bastante complexa. Eles devem ser definidos caso a caso, levando em conta o tipo dos dados que estão sendo tratados e suas características. Se o valor de épsilon escolhido for muito pequeno, um *cluster* (agrupamento) pode não ser formado devido à falta de pontos em uma região, fazendo com que a maioria dos pontos seja classificada como *outliers*. Em contrapartida, se o valor escolhido for excessivo, a maioria dos pontos do conjunto de dados será agrupada em um único *cluster*, resultando em um número reduzido de *outliers*. Desta forma, um valor pequeno é geralmente preferível. O mesmo ocorre com o número mínimo de pontos de uma amostra, se o valor for muito pequeno, ocorre uma maior propensão à formação de agrupamentos espúrios, ou seja, *clusters* compostos por ruídos ou pontos que não apresentam real similaridade. Por outro lado, se o número mínimo de pontos for elevado demais, pontos que deveriam formar clusters

distintos podem ser considerados ruídos ou agregados a *clusters* maiores, perdendo-se detalhes importantes na estrutura dos dados, especialmente em séries com maior variabilidade.

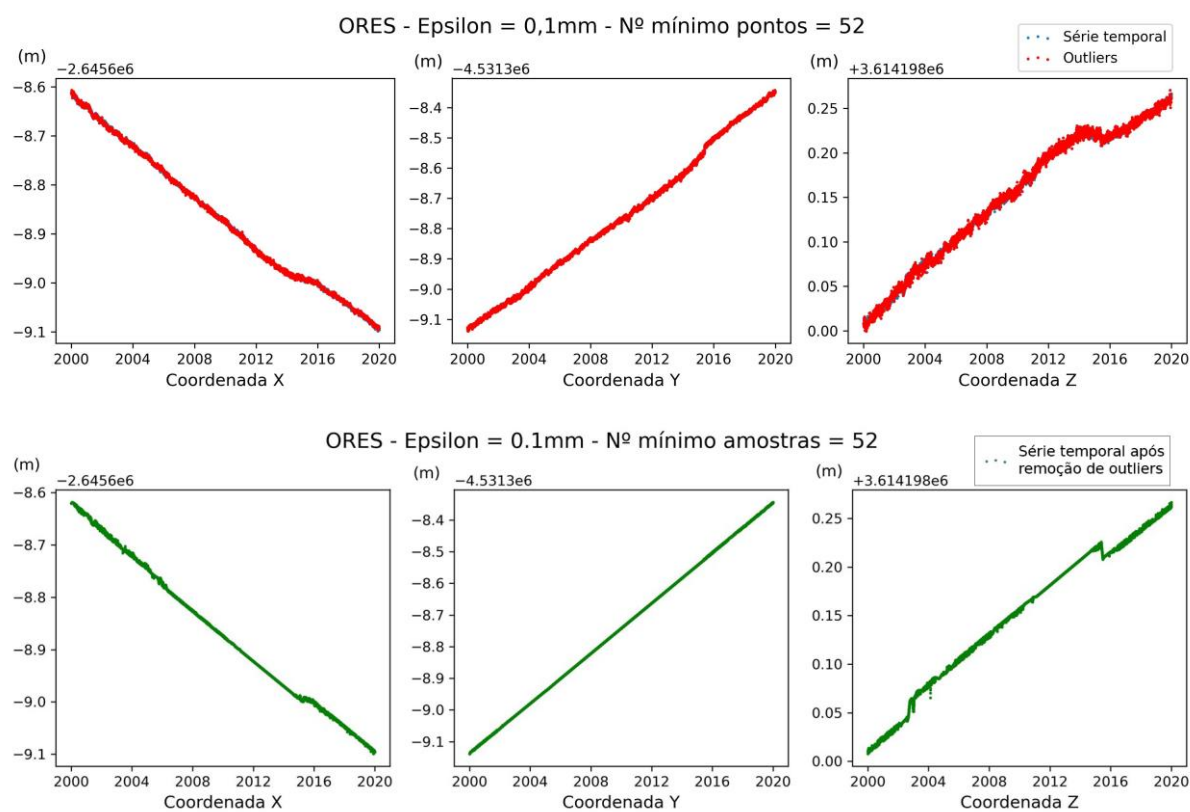
Considerando as 113 estações analisadas, a contaminação média foi de 16,27%. O valor máximo foi de 74,46%, registrado na estação ORES, enquanto o valor mínimo (0%) foi observado nas estações LDES e SHIN. Como destacam Nguyen e Tran (2023), uma contaminação de 15% é considerada alta, logo, contaminações como 61,55% na POMM, 72,62% na LAND, 61,95% na CAND, entre outras, não são proporções realistas de *outliers* em uma série temporal posicional do NGL. Contudo, o fato dessas altas contaminações proporcionarem discrepâncias menores apontam para uma linearização “forçada” das séries, ignorando as movimentações de curto período e reforçando a tendência de longo período das estações. Este comportamento pode ser observado nas estações OAES (Figura 48) e ORES (Figura 49), que apresentaram redução de discrepância equivalente a 0,001322 m e 0,001365 m, respectivamente.

Figura 48 – Séries da estação OAES com *outliers* identificados com DBSCAN e após remoção dos *outliers*.



Fonte: elaboração própria.

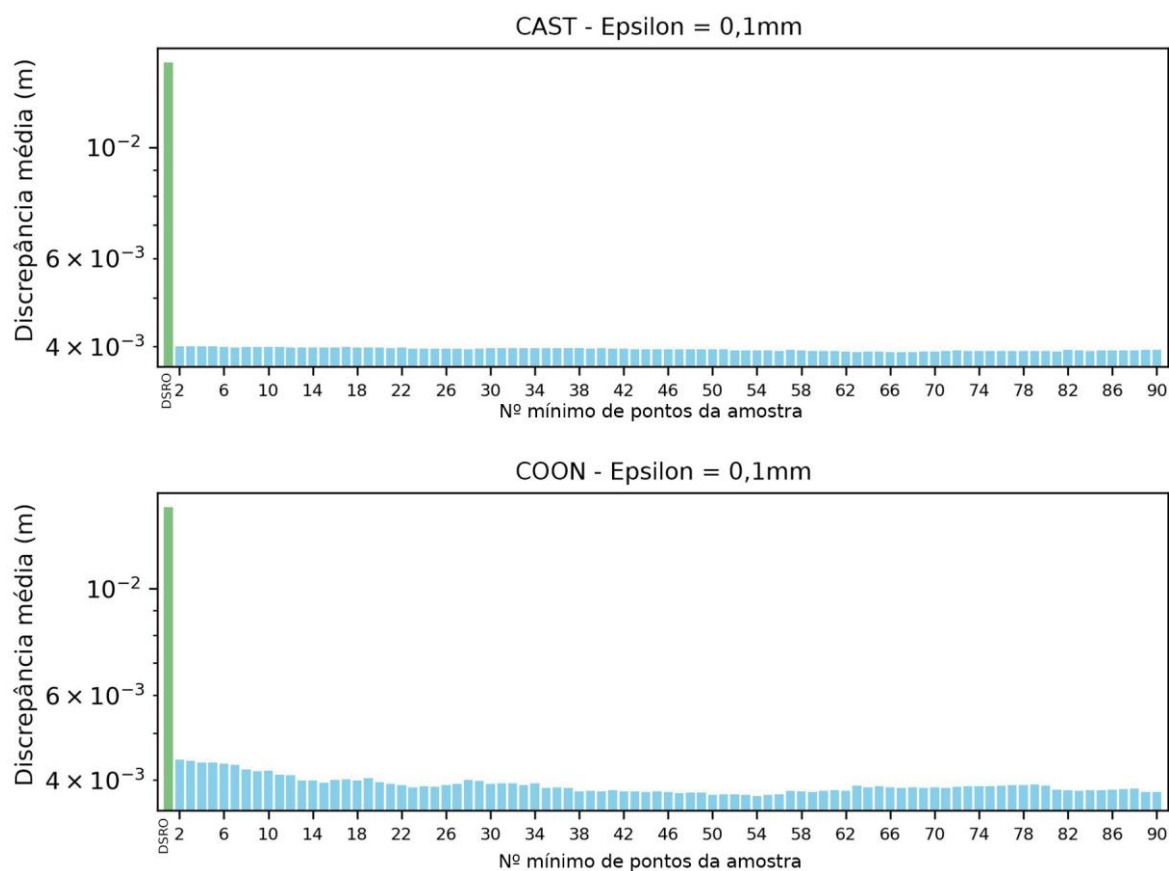
Figura 49 - Séries da estação ORES com *outliers* identificados com DBSCAN e após remoção dos *outliers*.



Fonte: elaboração própria.

Akbari e Unland (2016) verificaram que para dados 2D, não há uma variação significativa na detecção dos *outliers* com número mínimo de pontos da amostra maior que 4 (comportamento aqui verificado em diversas estações, como exemplificado na Figura 50), e, portanto, recomendam a utilização deste valor. O valor de 4 também é utilizado por Jain, Bajpai e Pamula (2022). Ademais, a função para aplicação do DBSCAN da biblioteca *Scikit learn* utiliza o valor 5 como padrão (Scikit-Learn, 2024b).

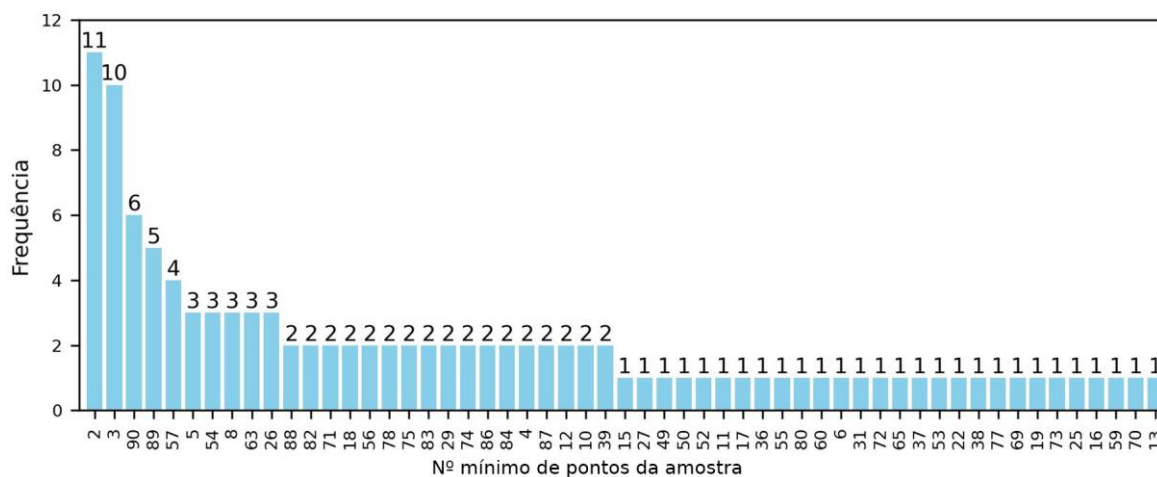
Figura 50 - Variação da discrepância média em relação ao valor do parâmetro “número mínimo de pontos da amostra” nas estações CAST e COON.



Fonte: elaboração própria.

Os valores do número mínimo de pontos da amostra que mais vezes proporcionaram a menor discrepância foram os valores de 2 e 3. Entretanto, estes valores foram os melhores em apenas 11 e 10 estações, respectivamente. Na sequência, os valores de 90, 89 e 57 apresentaram frequência de 6, 5 e 4, respectivamente. Já os demais valores, entre os que apresentaram a menor discrepância em alguma estação, ocorreram entre 1 e 3 vezes (Figura 51).

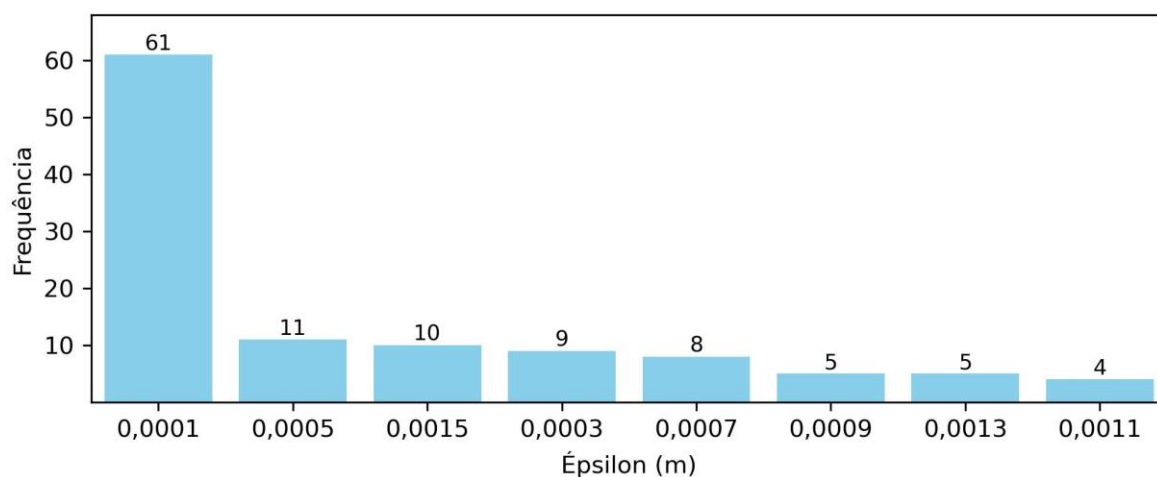
Figura 51 - Frequência do número mínimo de pontos da amostra que proporcionou a menor discrepância após aplicação do DBSCAN.



Fonte: elaboração própria.

Já o valor de ϵ com maior frequência foi 0,0001 m, em 61 estações (54% das estações), com os demais valores distribuídos em quantidades similares, como mostra a Figura 52.

Figura 52 - Frequência de ϵ que proporcionou a menor discrepância após aplicação do DBSCAN.



Fonte: elaboração própria.

De forma geral, os valores de ϵ de 0,0001 m e o do número mínimo de pontos entre 2 e 5 se mostraram os mais indicados para a aplicação do DBSCAN em séries temporais posicionais. Entretanto, destaca-se que dependendo das características intrínsecas das séries de cada estação, os resultados podem ser significativamente diversos.

6.4.9 Comparações dos métodos

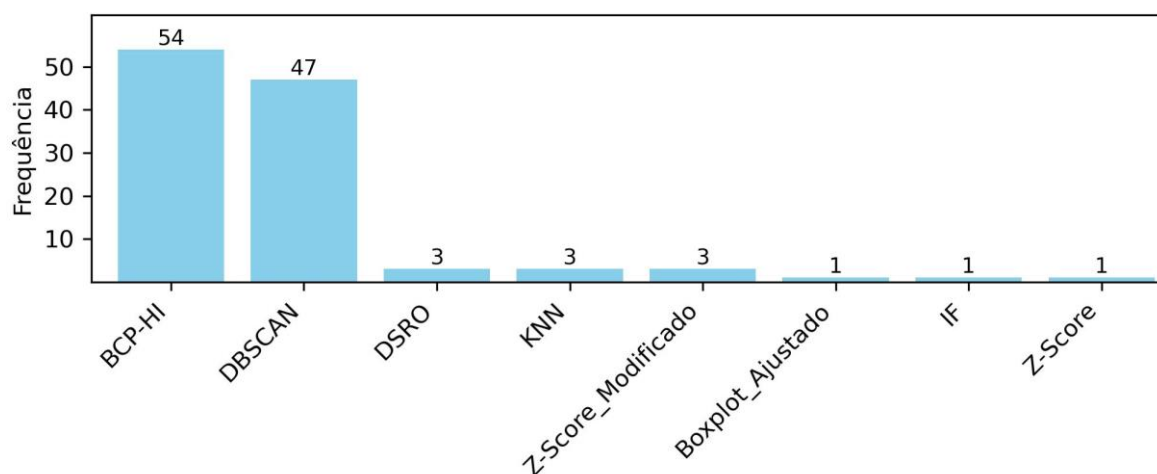
Todos os algoritmos de detecção de *outliers* possuem suas limitações, oriundas principalmente das suposições nas quais se baseiam. Diferentes famílias de algoritmos, como aqueles baseados em distribuição, distância, ângulo, densidade, entre outros, quando aplicados ao mesmo conjunto de dados, podem realizar decisões consideravelmente diferentes e variar amplamente na identificação de certos pontos dos dados e, portanto, em seus desempenhos (Papastefanopoulos; Linardatos; Kotsiantis, 2021). Somando-se a isso a variabilidade de características das estações, a performance dos métodos variou significativamente entre as séries.

A média móvel simples não se mostrou eficaz na eliminação de *outliers*, não alterando as velocidades estimadas, e conseqüentemente, as discrepâncias. Os métodos estatísticos (*Z-Score*, *Z-Score* modificado, *boxplot* ajustado e BCP-HI) se mostraram capazes de identificar *outliers* com certa precisão, mas com menor impacto na discrepância se comparado aos métodos de aprendizado de máquina (KNN, IF e DBSCAN). Tran (2013) destaca que normalmente, os métodos estatísticos são eficazes apenas quando a proporção de *outliers* é inferior a 5%.

Para *outliers* de maior magnitude, como os presentes na S300, todos os métodos realizaram uma identificação correta destes pontos. Em séries com maior variação de tendência de curto período, a maior diferença entre os métodos foi a restritividade dos limites além dos quais os pontos são classificados como *outliers*, com os métodos de aprendizado de máquina sendo os mais restritivos.

Ao verificar qual método proporcionou a menor discrepância para cada estação, nota-se que o BCP-HI aparece com a maior frequência, 54 vezes (Figura 53). Todavia, este método proporcionou uma redução da discrepância de mais de 0,001m em apenas uma estação, a S300, e aumentou a discrepância em 4 estações (Figura 54). Portanto, o BCP-HI apresentou uma performance consistente de redução de discrepância, mas abaixo de 0,001 m (média de 0,000229 m).

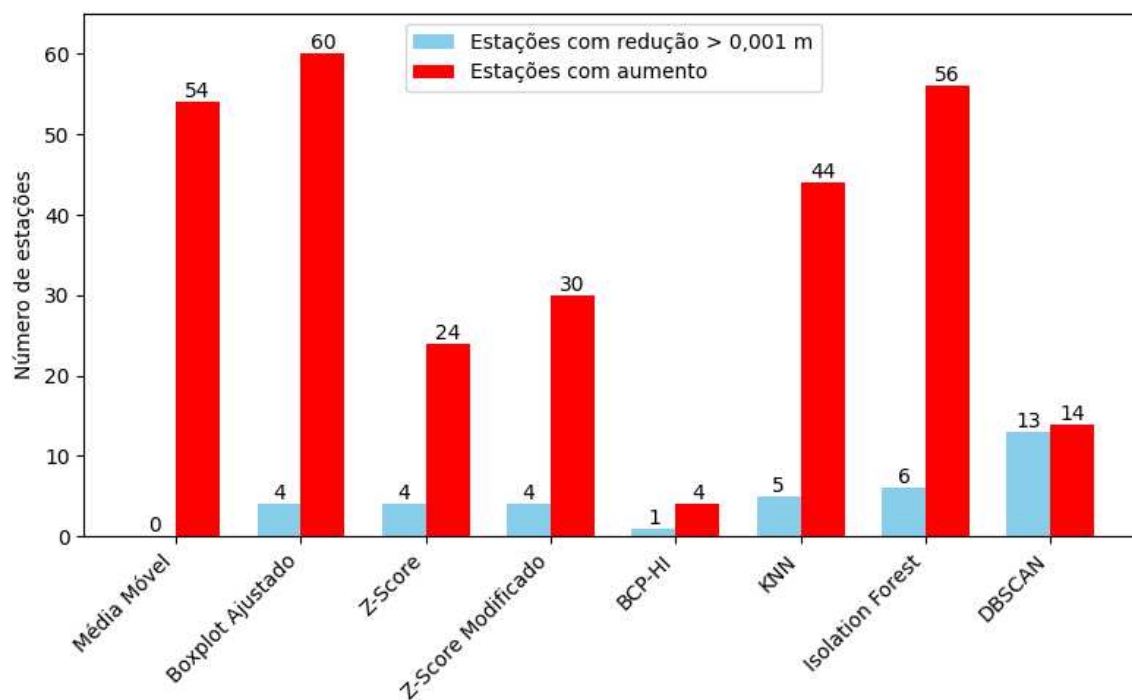
Figura 53 - Frequência em que cada método obteve a menor discrepância entre as 113 estações.



Fonte: elaboração própria.

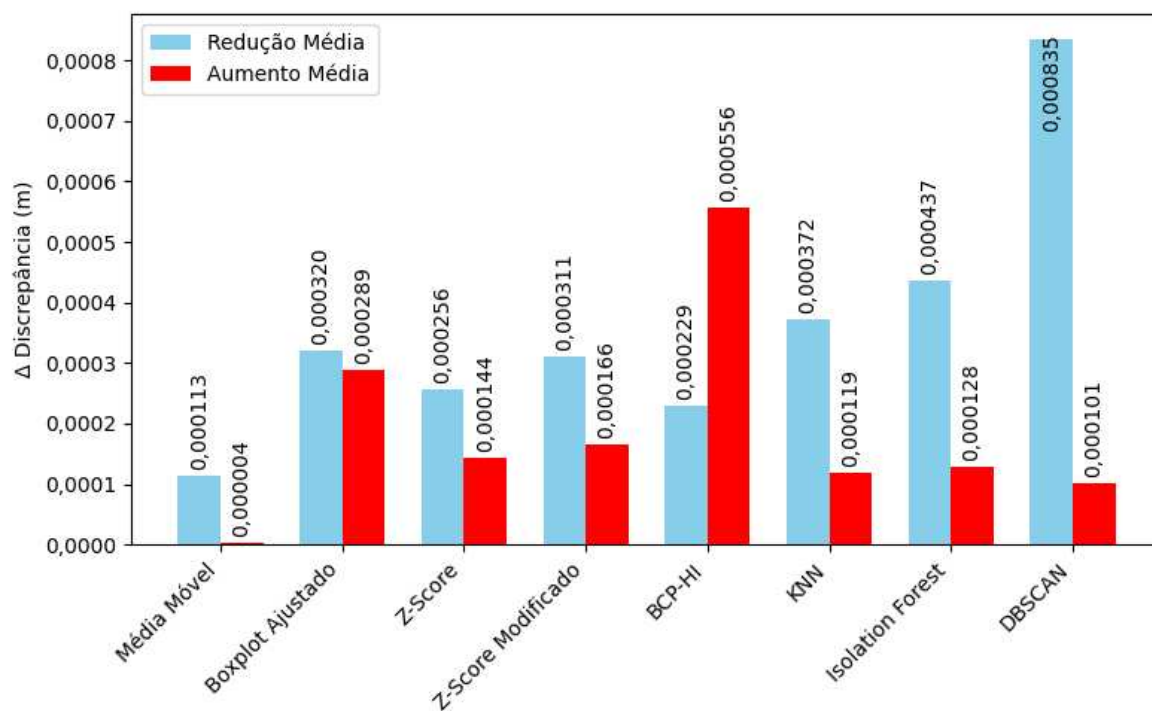
Em segundo lugar, aparece o DBSCAN, com melhor resultado em 47 estações, das quais, em 13, a redução foi acima de 0,001 m. Ademais, o DBSCAN apresentou a redução média mais alta entre os métodos testados, com 0,000835 m (Figura 55). Entretanto, Jain, Bajpai e Pamula (2022) destacam que convencionalmente, em séries temporais com componente sazonal, o DBSCAN funciona bem para detectar anomalias globais, mas falha em identificar as anomalias locais. Salienta-se também que o DBSCAN foi o método com a maior exigência computacional para ser aplicado.

Figura 54 - Número de estações em que a discrepância aumentou após a remoção de *outliers* e número de estações em que houve uma redução da discrepância superior à 0,001 mm.



Fonte: elaboração própria.

Figura 55 - Redução e aumento médio da discrepância planimétrica após remoção de *outliers*.



Fonte: elaboração própria.

Em 3 estações a remoção de *outliers* não melhorou a discrepância com nenhum dos métodos, sendo a DSRO a menor. O KNN e o *Z-Score* modificado foram melhores em 3 estações cada e o *boxplot* ajustado, o IF e o *Z-Score* foram melhores em apenas uma estação cada.

Analisando o valor da menor discrepância para cada uma das 113 estações entre todos os métodos, apenas 13 apresentaram redução de mais de 0,001 m após a remoção dos *outliers*, das quais, 11 foram obtidas com o DBSCAN, uma com o *Z-Score* e uma com o KNN, como mostra a Tabela 10.

Tabela 10 - Estações que apresentaram redução de discrepância acima de 0,001 m e o respectivo método.

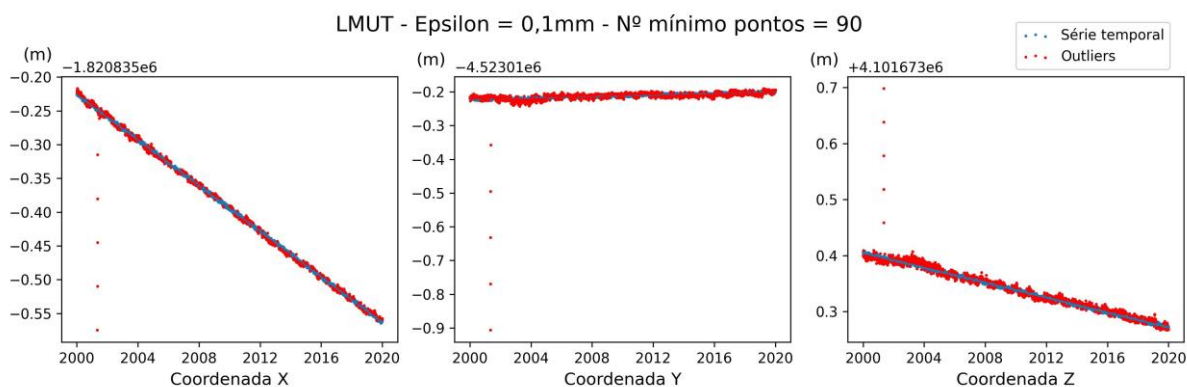
| Estação | Método com melhor discrepância | Δ Discrepância (m) |
|---------|--------------------------------|---------------------------------|
| CAND | DBSCAN | 0,011408 |
| CLAR | DBSCAN | 0,001016 |
| KNOL | DBSCAN | 0,001359 |
| KOKB | DBSCAN | 0,005565 |
| LAND | DBSCAN | 0,012505 |
| MIDA | DBSCAN | 0,010400 |
| OAES | DBSCAN | 0,001322 |
| OPCP | DBSCAN | 0,001076 |
| ORES | DBSCAN | 0,001365 |
| POMM | DBSCAN | 0,007755 |
| S300 | <i>Z-Score</i> | 0,007753 |
| TOST | KNN | 0,001871 |
| WMAP | DBSCAN | 0,005627 |

Fonte: elaboração própria.

Como descrito ao longo de todos os métodos, todas as 13 estações descritas na Tabela 10 possuem problemas relevantes em suas séries, como trocas de equipamentos, terremotos ou *outliers* de grande magnitude não relacionados a tais eventos, como a S300. Diante disso, pode-se inferir que para as séries que não se enquadrem nesses cenários, a presença de *outliers* de magnitude milimétrica ou até centimétrica não exerce uma influência significativa na velocidade estimada de suas séries, e conseqüentemente, em sua discrepância. Isso é evidenciado na LMUT, que possui 5 *outliers* com magnitude variando de 0,100 m a mais de 0,600 m na coordenada Y, e que nenhum método conseguiu reduzir sua discrepância após corretamente identificar e remover estes *outliers*. De fato, a menor discrepância para esta estação não foi obtida com nenhum dos métodos testados, e sim com a série original, sem

remoção dos *outliers*. A Figura 56 mostra as séries da LMUT, com os *outliers* identificados com DBSCAN.

Figura 56 - Séries da estação LMUT com *outliers* identificados com DBSCAN.



Fonte: elaboração própria.

Desse modo, velocidades estimadas a partir de séries posicionais de longa duração não mostraram alta sensibilidade a *outliers*. Por outro lado, na presença de saltos, os métodos de remoção de *outliers* foram capazes de melhorar as discrepâncias, mesmo que de forma pouco significativa.

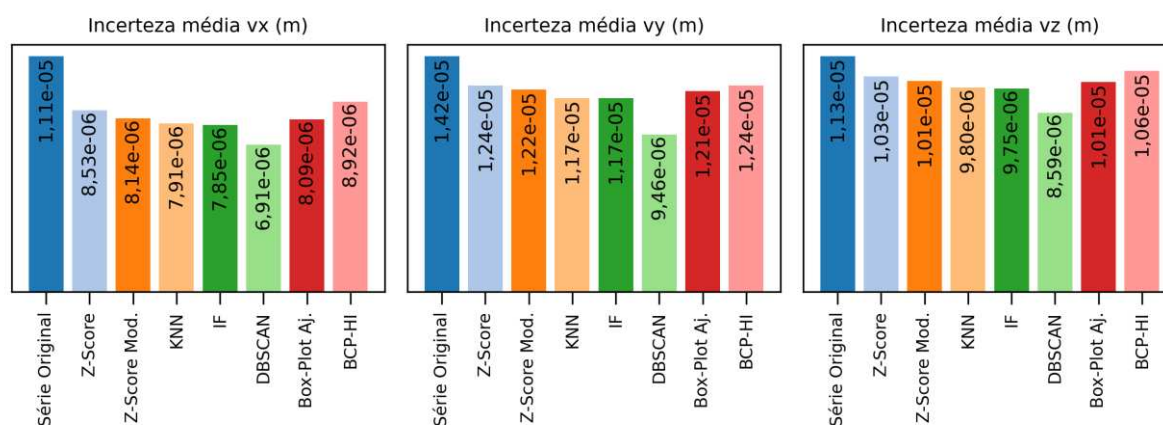
Masson, Mazzotti e Vernant (2019), que analisaram séries temporais posicionais sintéticas, verificaram que um dos principais contribuintes para a variabilidade nas velocidades estimadas de séries entre 8 e 21 anos é a presença de saltos. Portanto, a correção de saltos, especialmente os devidos a trocas de equipamentos, devem ser priorizados em relação à detecção de *outliers*.

Vale destacar que os métodos de detecção inevitavelmente irão identificar alguns pontos de dados como *outliers* quando não o são - assim como podem falhar em identificar *outliers* verdadeiros. Desta forma, a eliminação desses pontos falso positivos, pode influenciar negativamente a discrepância. Um cenário em que vários métodos identificaram erroneamente *outliers* foi em saltos provocados por terremotos e as respectivas acomodações pós-sísmicas, em que as séries apresentam um comportamento exponencial, e não mais linear, como exemplificado pela estação MIDA na Figura 44. Uma forma de lidar com essas deformações pós-sísmicas (*Post-Seismic Deformation - PSD*) para estações que foram afetadas por terremotos de grande magnitude é utilizando modelos de PSD - introduzidos no ITRF2014 (Altamimi *et al.*, 2016) e aperfeiçoados no ITR2020 (Altamimi *et al.*, 2023) -, estimados após

o ajuste de funções exponenciais e/ou logarítmicas aos dados GNSS. Salienta-se que a aplicação destes modelos não necessariamente implica em uma mudança no valor da velocidade (Altamimi *et al.*, 2023).

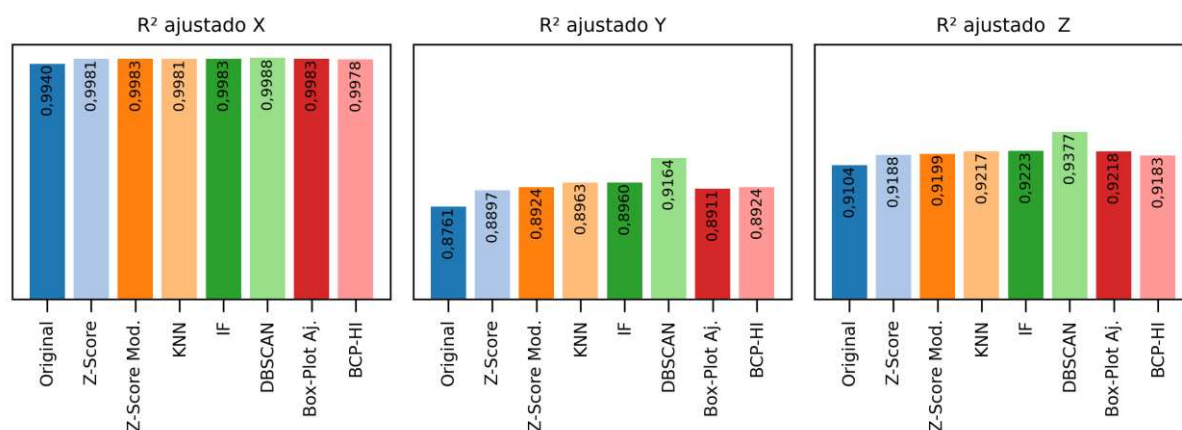
Em relação à incerteza das velocidades, é possível notar que a remoção dos *outliers* com todos os métodos resultou em uma redução destes valores, como mostrado na Figura 57. Entretanto, deve-se destacar que as incertezas já são submilimétricas nas séries originais, portanto essa diminuição não é muito significativa. Assim como para as discrepâncias, o DBSCAN também mostrou os melhores resultados em relação aos demais métodos para a redução de incertezas.

Figura 57 - Incerteza média das velocidades estimadas a partir das séries originais e das séries após remoção de *outliers*.



Fonte: elaboração própria.

O R^2 ajustado médio das 113 estações para as coordenadas X, Y e Z estão dispostos na Figura 58. Um R^2 ajustado igual a 1 significa que o modelo explica 100% da variação da variável dependente, ou seja, indica um ajuste perfeito (Montgomery; Peck; Vining, 2021). Para as séries da coordenada X, temos valores de R^2 ajustado de 0,99, tanto antes como depois da remoção de *outliers*, sinalizando uma grande linearidade destas séries. Para as coordenadas Y, temos valores de R^2 ajustado menores, sendo estas séries as menos lineares entre as 3 componentes. A remoção de *outliers* com todos os modelos aumentara o R^2 ajustado, mas o DBSCAN se destacou, elevando de 0,876 da série original para 0,916 depois da remoção de *outliers*. Para as coordenadas Z, o comportamento foi similar ao das coordenadas Y, com o DBSCAN também proporcionando o maior aumento no valor do R^2 ajustado.

Figura 58 - O R^2 ajustado médio das 113 estações para as coordenadas X, Y e Z.

Fonte: elaboração própria.

No que diz respeito ao teste do qui-quadrado, em todas as séries temos $X^2 = 7303$, que é igual a $n - p - 1$ ($7305 - 1 - 1$), o que, como indica Wooldridge (2016), significa que os resíduos estão alinhados com as expectativas de um modelo bem ajustado e a distribuição dos erros é consistente com os pressupostos.

Para o p -valor, também foi encontrado um valor semelhante para todas as séries (diferença da ordem de 10^{-14}), p -valor = 0,497799335594891. Como o valor está próximo de 0,5, isso indica que o valor do qui-quadrado está alinhado com os graus de liberdade, reforçando que os modelos estão adequados (Agresti, 2013).

6.4.10 Remoção de tendência e sazonalidade

Após a remoção de tendência e sazonalidade das séries, a aplicação dos métodos de identificação e remoção de *outliers* produziu resultados menos satisfatórios, se comparados àqueles obtidos com a remoção apenas da tendência. De forma geral, foi identificado um menor número de *outliers*, o que, conseqüentemente, resultou em um menor impacto na discrepância. Para todos os métodos, apenas na estação S300 a remoção de *outliers* reduziu a discrepância em mais de 0,001 m. A disparidade entre os resultados pode ser observada na Figura 59.

Figura 59 - Comparação do número de estações com redução de discrepância maior que 0,001 m após a remoção de *outliers* em séries sem tendência e sem tendência e sazonalidade.



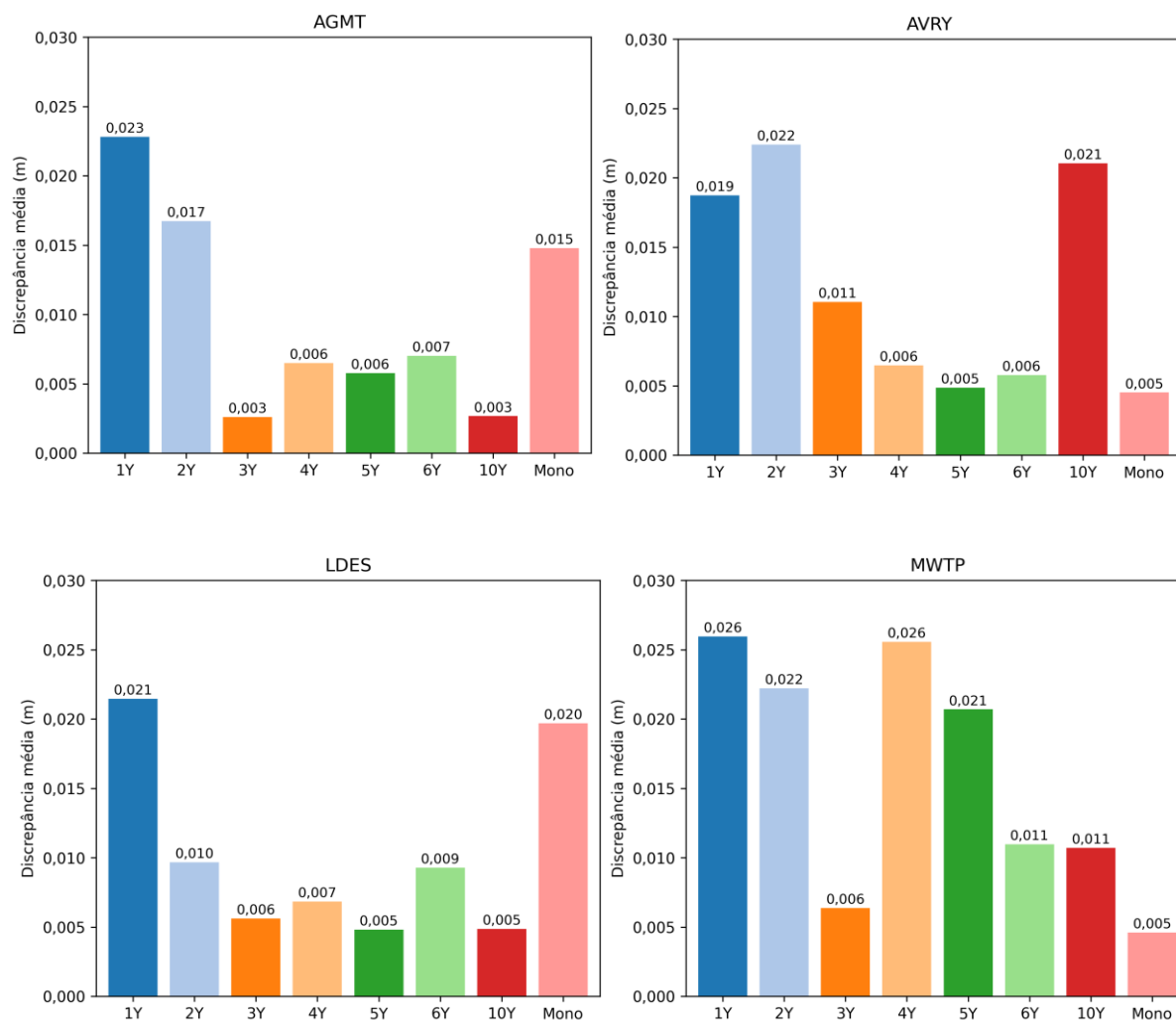
Fonte: elaboração própria.

Desta forma, optou-se por não analisar os experimentos realizados nas séries com a remoção de tendência e sazonalidade. Entretanto, as tabelas completas para as 113 estações, empregando todos os 6 métodos estão disponíveis em [bit.ly/Rem Tend e Saz](http://bit.ly/Rem_Tend_e_Saz).

6.5 Multivelocidade

Calculadas as discrepâncias planimétricas para as 113 estações utilizando a monovelocidade de 20 anos (mono) e as multivelocidade de 1,2,3,4,5,6 e 10 anos (1Y, 2Y, 3Y, 4Y, 5Y, 6Y e 10Y, respectivamente), partiu-se para a exploração dos resultados. Analisando algumas estações individualmente, nota-se grande variabilidade na performance das velocidades, como exemplificado pela Figura 60.

Figura 60 - Discrepância planimétrica média para as estações AGMT, AVRY, LDES e MWTP para cada velocidade.



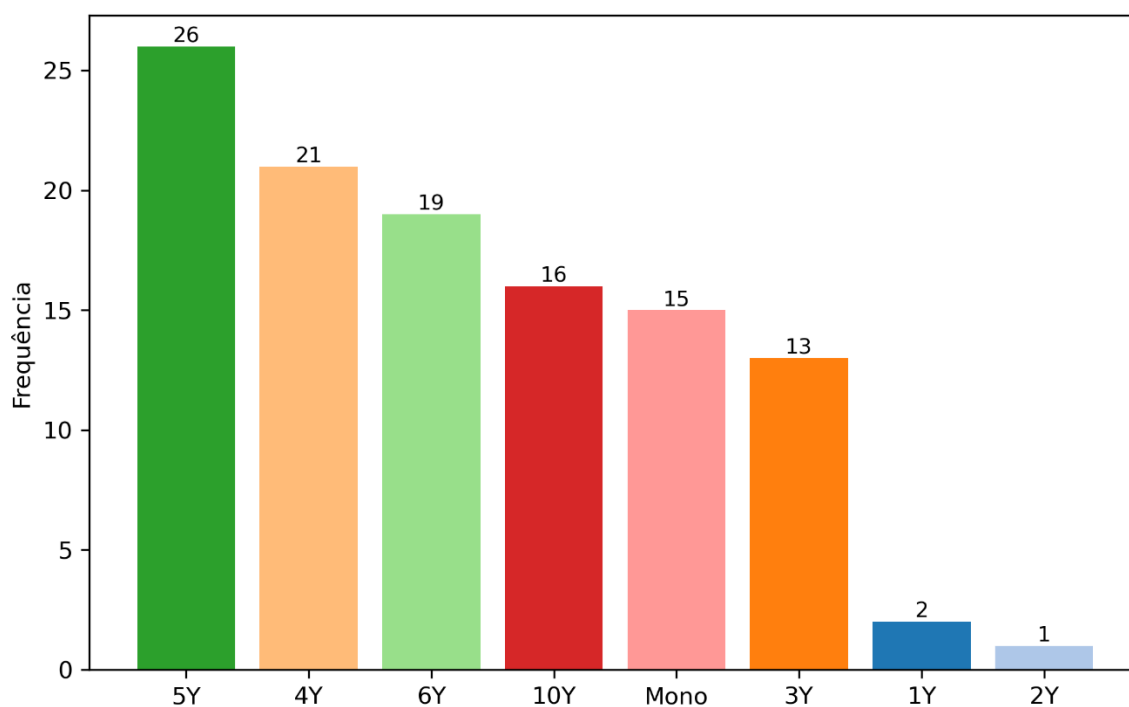
Fonte: elaboração própria.

Na AGMT, a discrepância com a 20Y foi de 0,015 m, enquanto com as multivelocidades 3Y e 10Y foi de 0,003m. Na AVRY, a mono e a 5Y apresentaram discrepância de 0,005 m, contra 0,021 m da 10Y. Na LDES a 20Y só foi menor que a 1Y, com discrepância de 0,020 m. As melhores multivelocidades foram as de 5Y e 10Y, com discrepância de 0,005 m.

Já na MWTP, a menor discrepância foi obtida com a mono (0,005 m), seguida pela 3Y (0,006 m). Entretanto, as demais multivelocidades foram significativamente piores, variando entre 0,011 e 0,026 m. A 4Y apresentou resultado semelhante ao da 1Y, e a 5Y semelhante ao da 2Y.

Visando uma análise mais geral das metodologias, identificou-se a menor discrepância obtida entre todas as multivelocidades e a monovelocidade para cada uma das 113 estações, a frequência de cada velocidade é exibida na Figura 61.

Figura 61 - Frequência da velocidade que resultou na melhor discrepância planimétrica calculada para as 113 estações.



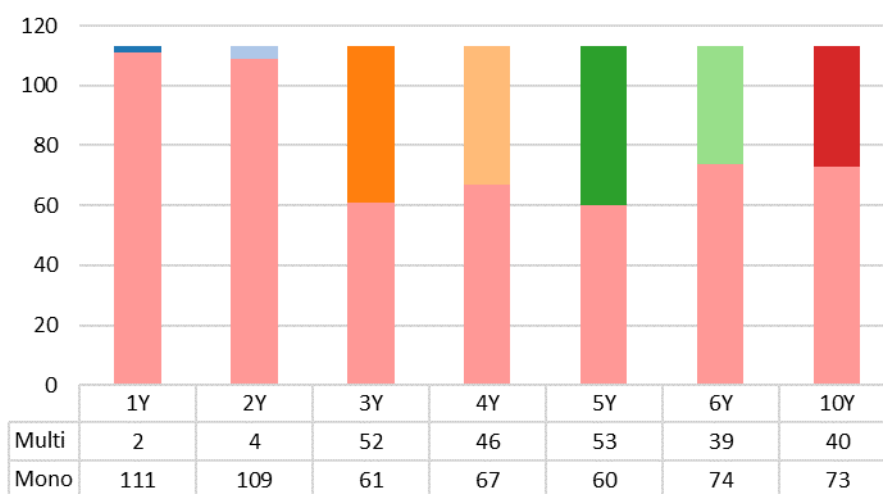
Fonte: elaboração própria.

A 5Y apresentou o maior número de menores discrepâncias, em 26 estações (23%), seguida pela 4Y, 6Y e 10Y, com menores discrepâncias em 21, 19 e 16 estações, respectivamente. A mono aparece apenas na quinta posição, tendo sido melhor em 15 estações (13%), superando a 3Y, melhor em 13 estações.

As 1Y e 2Y se destacaram negativamente, sendo melhores em apenas 2 e 1 estações, respectivamente. De fato, SIRGAS (2024b) destaca que velocidades devem ser preferencialmente estimadas através de séries temporais posicionais com período mínimo de dois anos. Wang (2021) recomenda observações contínuas por mais de 2,5 e 4 anos para alcançar um intervalo de confiança de 95% abaixo de 1 mm/ano e 0,5 mm/ano, respectivamente. Nesse contexto, as velocidades de curto período são mais sensíveis à sazonalidade das séries, como destacam Blewitt *et al.* (2016), assim como a saltos e *outliers*.

Comparou-se também cada multivelocidade isoladamente com a monovelocidade, registrando qual apresentou a menor discrepância em cada estação (Figura 62). Mais uma vez fica evidente a pior performance das 1Y e 2Y, que foram melhores que a mono em apenas 2 e 4 estações, respectivamente. As 3Y e 5Y foram as que superaram a mono no maior número de estações, 52 (46%) e 53 (47%) respectivamente.

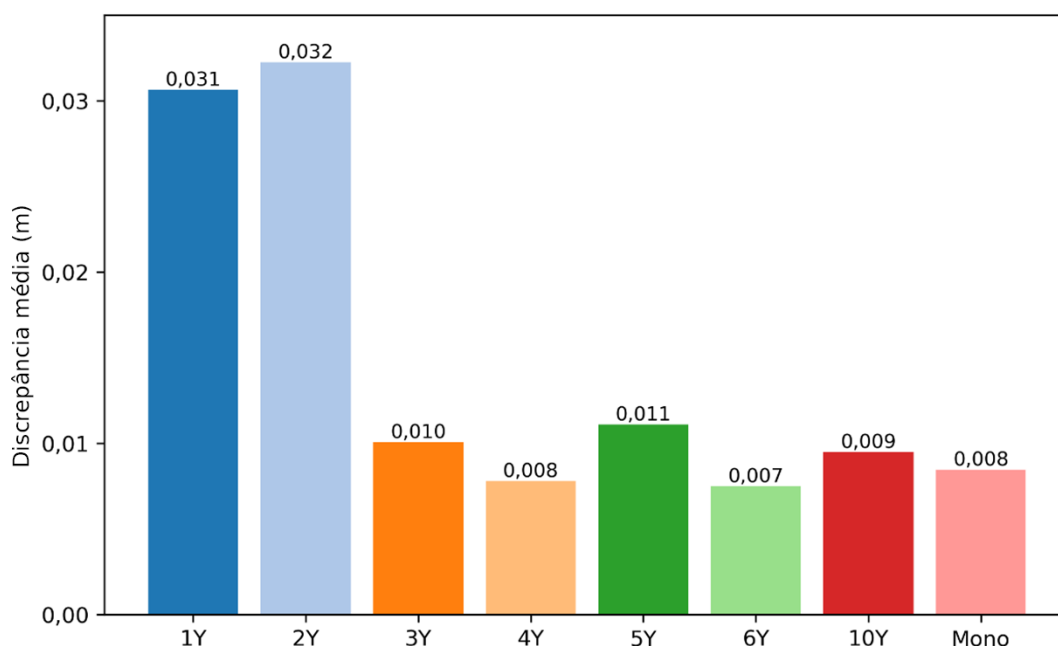
Figura 62 – Frequência de menor discrepância planimétrica na comparação de cada multivelocidade isoladamente com a monovelocidade (rosa) para as 113 estações.



Fonte: elaboração própria.

Analisando a discrepância média das 113 estações para cada velocidade (Figura 63), novamente nota-se os piores resultados sendo das 1Y e 2Y. Entre as demais, a diferença entre a pior, 5Y, com discrepância de 0,011 m, e a melhor, 6Y, com discrepância de 0,007 m, foi de 0,004 m. Ambas bem próximas da discrepância da monovelocidade, que foi de 0,008 m.

Figura 63 - Discrepância planimétrica média das 113 estações para cada velocidade.



Fonte: elaboração própria.

Masson, Mazzotti e Vernant (2019), através de análises de árvores de regressão de séries sintéticas, mostraram que os principais fatores que controlam a precisão das velocidades, em ordem de importância, são: o período da série, a presença de saltos e o nível de ruído (dispersão e índice espectral). Ademais, concluem:

- séries com período inferior a 4,5 anos, com presença de saltos, não devem ser utilizadas para aplicações que exigem uma precisão superior que 1,0 mm/ano;
- séries entre 4,5 e 8 anos, com ausência de saltos, estão associadas a um baixo viés ($v_{95} \leq 0,3$ mm/ano, onde v_{95} é o quantil de 95% da distribuição de viés, representando 95% de confiança nas velocidades estimadas). As séries que possuem pelo menos um salto estão associadas a um viés de $v_{95} \approx 0,6$ mm/ano;
- os valores de viés de velocidade mais alto, devido apenas ao ruído, podem alcançar $v_{95} \leq 0,7$ mm/ano para séries curtas (menores que 5 anos); e,
- para séries com mais de 15 anos, todos os v_{95} são menores que 0,1 mm/ano.

Diante disso, pode-se verificar algumas vantagens e desvantagens das multivelocidades: por um lado, se adequam melhor às tendências de curta duração das séries, representando melhor seus deslocamentos; por outro, são afetadas pelos problemas relacionados às séries

curtas (amplificadas em séries com menos de 3 anos), como menor precisão na estimativa de velocidades e maior sensibilidade a saltos, ruídos e sazonalidade.

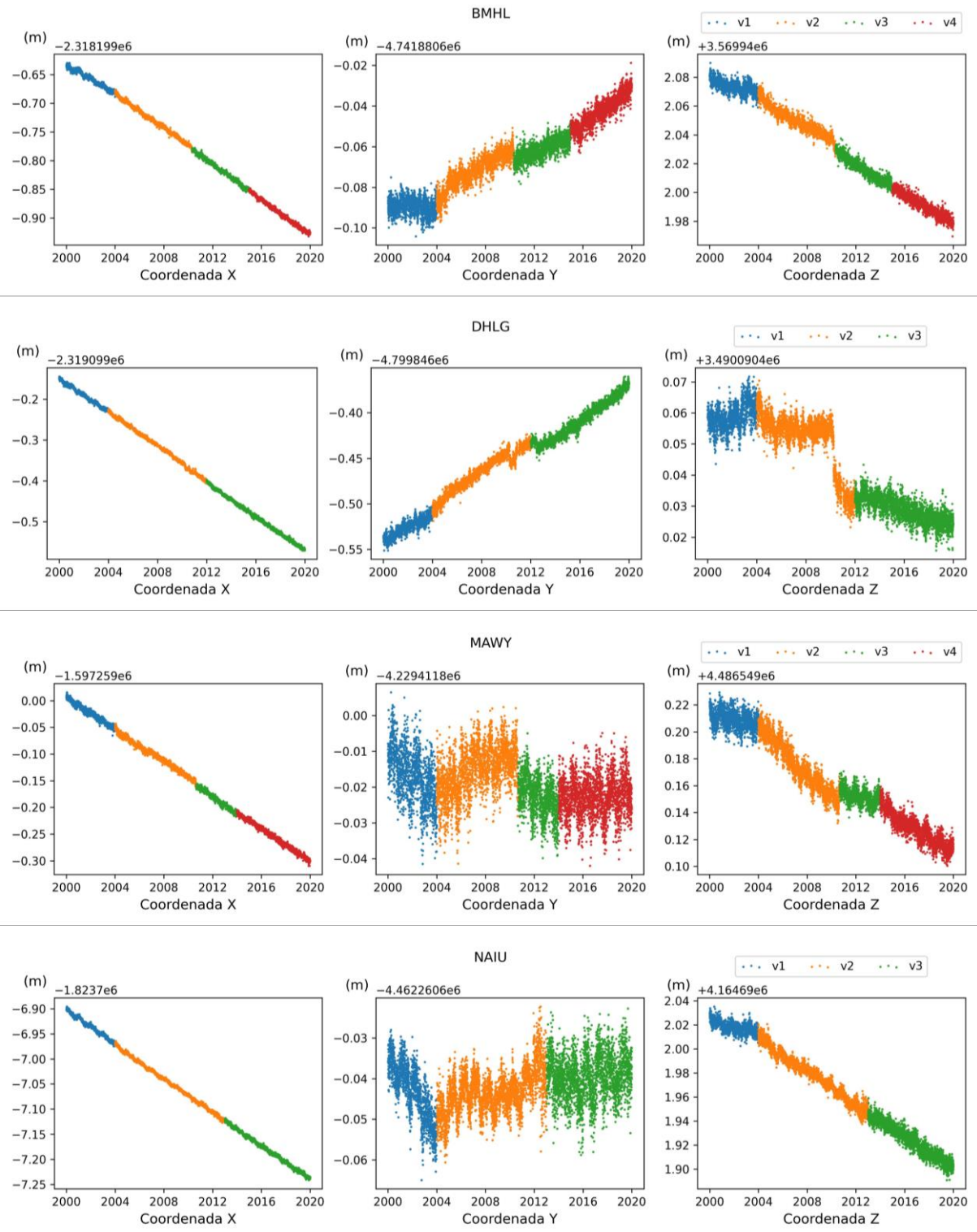
Freitas, Dal Poz e Nascimento (2022), que analisaram multivelocidades de 2, 4 e 5 anos e monovelocidade de 8 e 10 anos, para 3 estações do Brasil e 3 estações dos Estados Unidos, constataram que as multivelocidades de 4 e 5 anos apresentaram resultados semelhantes ou superiores àqueles da monovelocidade.

De forma geral, considerando os resultados das 113 estações de forma conjunta, temos que as multivelocidades entre 3Y e 10Y apresentaram discrepâncias médias semelhantes àqueles da monovelocidade. Comparando-se todas as multivelocidades com a monovelocidade, temos que a mono foi melhor em apenas 13% das estações. Entretanto, na comparação isolada, temos que as multivelocidades apresentaram a menor discrepância em no máximo 47% das estações. Por outro lado, analisando cada estação isoladamente os resultados são muito diversos, não sendo possível prever nem quando ou qual multivelocidade será melhor que a monovelocidade.

Uma vez que os resultados com as multivelocidades com períodos uniformes não foram satisfatórios, partiu-se para a análise de períodos distintos para uma mesma série, de forma a adequar melhor cada velocidade ao comportamento de cada série. Foi observado um período mínimo de 3 anos para cada segmento, uma vez que os resultados com 1 e 2 anos apresentaram uma degradação da discrepância após a atualização.

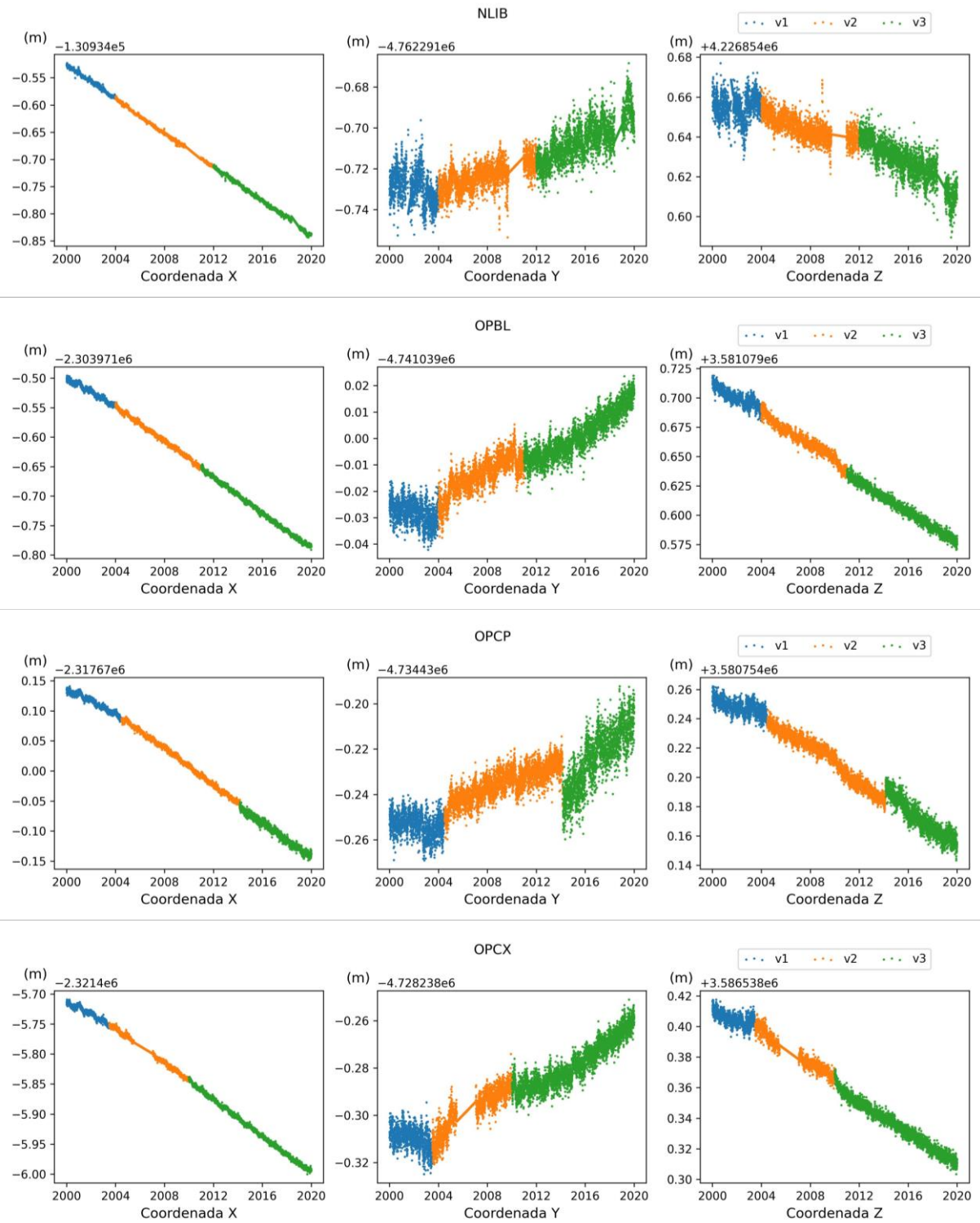
Foram selecionadas 10 estações que apresentaram um R^2 ajustado baixo, indicando uma série menos linear, sendo elas: BMHL, DHLG, MAWY, NAIU, NLIB, OPBL, OPCP, OPCX, OPRD e SLMS. Sete das séries foram segmentadas em 3 partes e 3 séries foram segmentadas em 4 partes, de acordo com suas características de deslocamento e mudanças de tendências (Figura 64, Figura 65 e Figura 66). Uma tabela com as informações dos segmentos está disponível no APÊNDICE B.

Figura 64 - Segmentação das séries das estações BMHL, DHLG, MAWY e NAIU.



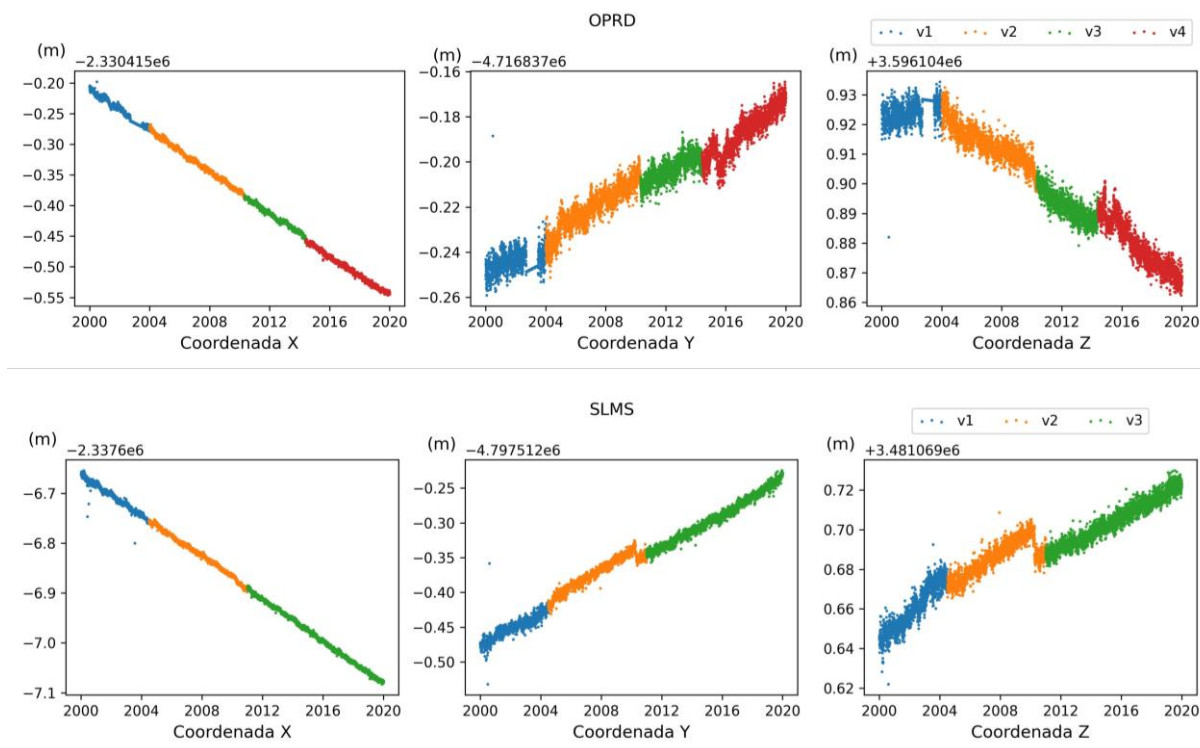
Fonte: elaboração própria.

Figura 65 - Segmentação das séries das estações NLIB, OPBL, OPCP e OPCX.



Fonte: elaboração própria.

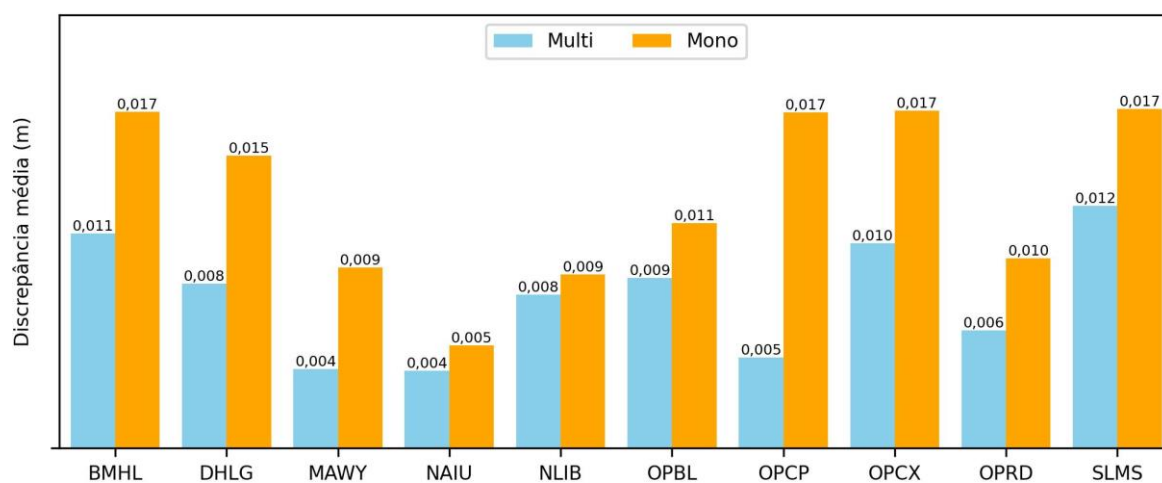
Figura 66 - Segmentação das séries das estações OPRD e SLMS.



Fonte: elaboração própria.

Após a segmentação das séries, estimativa das velocidades e atualização das mesmas 365 coordenadas do ano de 2019 para a época 2000.0, calculou-se as discrepâncias planimétricas médias e as comparou com aquelas calculadas com a monovelocidade. Os resultados podem ser visualizados na Figura 67.

Figura 67 - Discrepâncias médias calculadas após atualização de coordenadas com mono e multivelocidades.

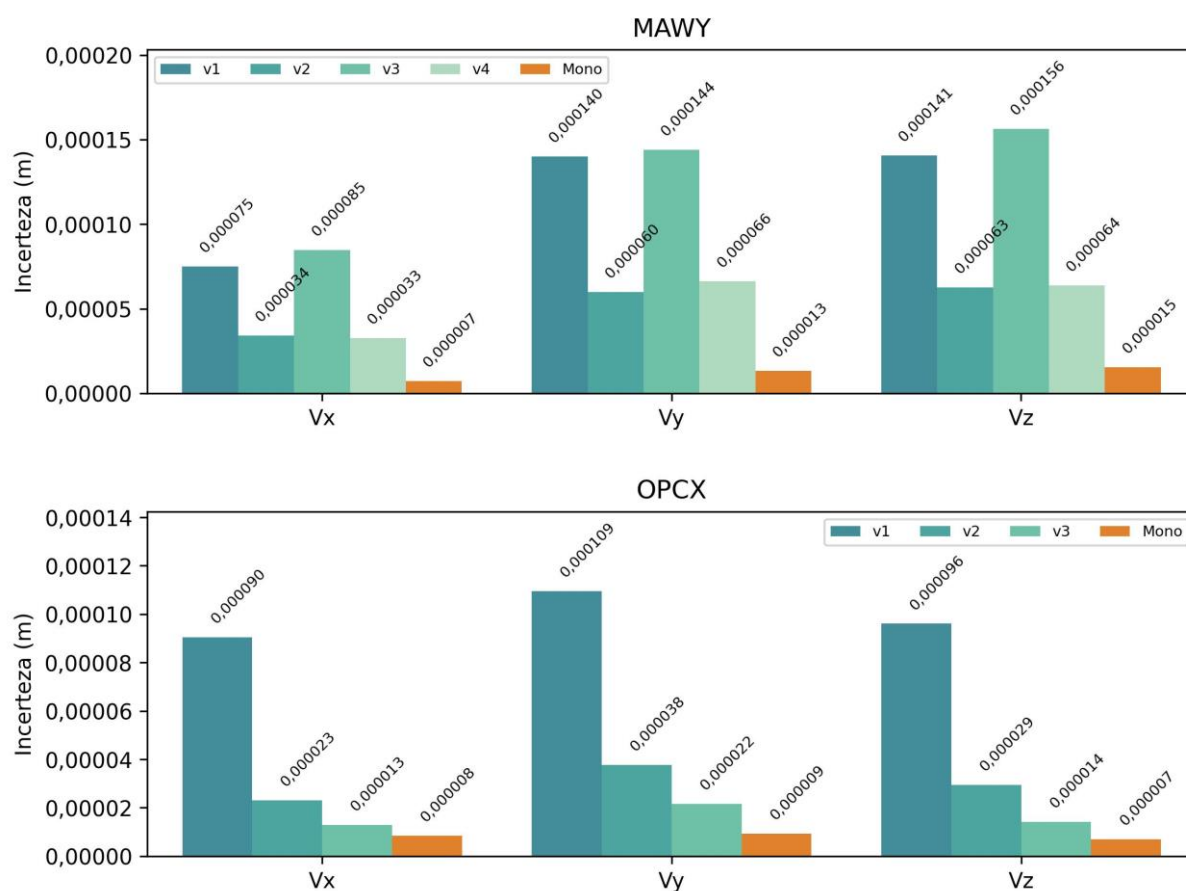


Fonte: elaboração própria.

Em todas as estações a multivelocidade proporcionou uma discrepância menor que a monovelocidade, mostrando a vantagem de adequar os segmentos das séries de forma a melhor representar os deslocamentos, principalmente quando há mudanças mais relevantes da tendência de longa duração. Isso é mais bem exemplificado pela estação OPCP (Figura 65), em que se pode observar 3 tendências distintas na série de coordenadas Y, o que proporcionou a maior redução de discrepância em relação à monovelocidade, de 0,012 m.

Uma preocupação ao se realizar vários processos de atualização de coordenadas seguidos, como nas multivelocidades, consiste na degradação das incertezas das coordenadas finais. Aliado a isto, deve-se notar também que por serem estimadas a partir de séries mais curtas, as multivelocidades já possuem incertezas maiores em relação as monovelocidades, como exibido na Figura 68.

Figura 68 - Incertezas das multivelocity das estações MAWY e OPCX.

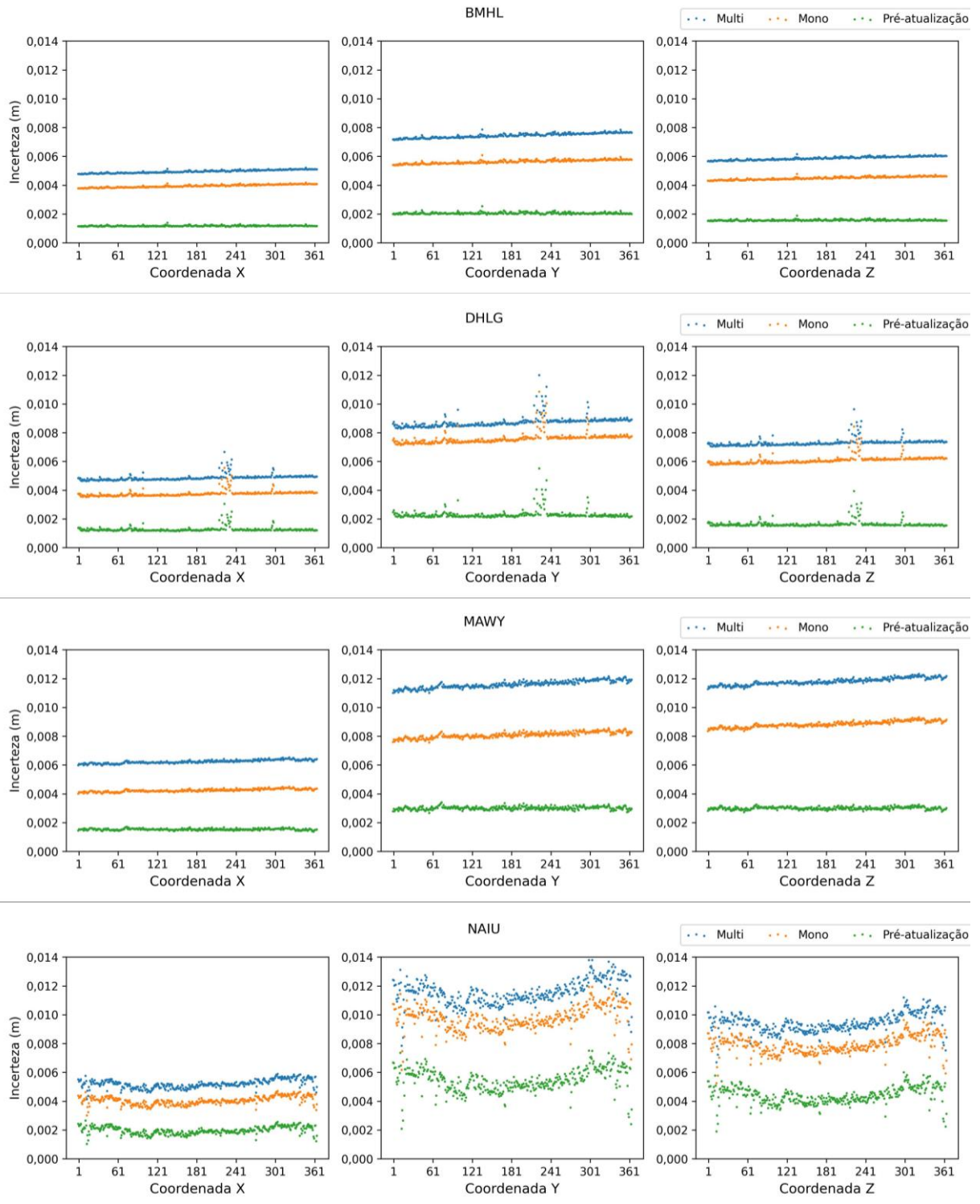


Fonte: elaboração própria.

De fato, o tamanho dos segmentos tem impacto direto nas incertezas das velocidades, como pode ser observado na MAWY, em que o primeiro e o terceiro segmento são os de menor período, e conseqüentemente, suas respectivas velocidades apresentam as maiores incertezas. Na OPCX, em que o primeiro segmento é significativamente menor que o segundo e o terceiro, verifica-se também uma diferença nas incertezas.

As incertezas propagadas das coordenadas, após as atualizações, são exibidas nas Figura 69. Destaca-se que o processo de propagação foi repetido a cada etapa intermediária de atualização entre a época inicial e a final de atualização. Isto é, como a atualização de 2019 para 2000 empregou 3 ou 4 atualizações (de acordo com o número de segmentos que a série foi dividida), realizou-se uma propagação para cada atualização.

Figura 69 - Incertezas das coordenadas atualizas com multivelocidade das estações BMHL, DHLG, MAWY e NAIU.



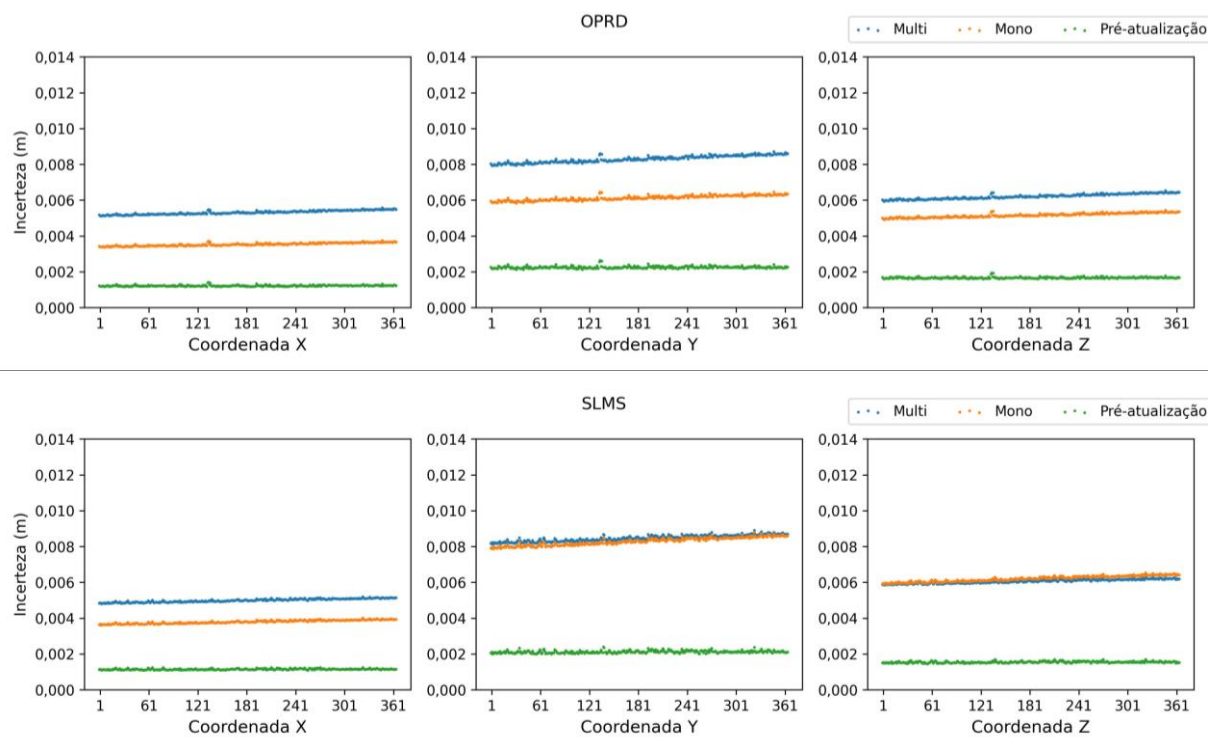
Fonte: elaboração própria.

Figura 70 - Incertezas das coordenadas atualizadas com multivelocidade das estações NLIB, OPBL, OPCP e OPCX.



Fonte: elaboração própria.

Figura 71 - Incertezas das coordenadas atualizadas com multivelocidade das estações OPDR e SLMS.



Fonte: elaboração própria.

De acordo com os resultados, não foi possível verificar uma diferença significativa entre o uso de 3 ou 4 multivelocidades, tanto na discrepância quanto nas incertezas das velocidades. Estações como a NLIB e a NAIU, que foram segmentadas em 3 partes apresentaram incertezas maiores que a BMHL e a OPDR que foram divididas em 4 partes.

De forma geral, a atualização com multivelocidades aumentou as incertezas entre 1 e 3 mm, comparado aos valores da atualização com monovelocidade. Desta forma, com a redução da discrepância e uma degradação não significativa das coordenadas atualizadas, a atualização por multivelocidade se mostrou um método eficaz para a atualização de coordenadas de estações com séries que apresentam saltos ou mudanças de tendência.

7 CONCLUSÕES

Empregando séries temporais de dados reais, que refletem a complexidade e a variação natural intrínseca da natureza, verificou-se que a remoção de *outliers* das séries não apresentou impacto significativo nas discrepâncias planimétricas finais. Ou seja, as velocidades não se mostraram sensíveis aos *outliers* das séries das quais são estimadas. Destaca-se que as análises

realizadas consideram o processo de atualização de coordenadas geodésicas que será utilizado majoritariamente para fins de posicionamento e trabalhos cartográficos; neste contexto, a remoção de *outliers* só se mostrou necessária para *outliers* de grande magnitude e em grande número.

Quanto aos métodos de imputação de dados ausentes, verificou-se que a interpolação linear se mostrou consistente e proporcionou menores discrepâncias após a atualização de coordenadas.

Em relação aos métodos de estimativa de velocidades, verificou-se que os métodos baseados nos mínimos quadrados são intercambiáveis, uma vez que seus resultados mostraram diferenças desprezíveis para esta finalidade. As velocidades estimadas com a regressão linear simples e o MIDAS também foram similares para a maioria das estações. Entretanto, destaca-se que o MIDAS apresentou resultados significativamente piores em algumas séries com saltos.

Mesmo sem um pré-processamento das séries e analisando regiões de alta atividade sísmica (há uma alta concentração de estações no oeste dos EUA, próximas ao encontro das placas litosféricas do Pacífico e Norte-americana) foi possível atualizar coordenadas por 20 anos com baixas discrepâncias planimétricas, demonstrando a eficiência e eficácia desta metodologia para estimativa de velocidades, tanto pela facilidade de execução, baixo custo computacional e qualidade dos resultados.

Diante da facilidade de obtenção das séries temporais fornecidas pelo NGL utilizando o GPTS Downloader e da simplicidade da estimativa das velocidades, o fator limitante do emprego da metodologia analisada consiste na disponibilidade de uma estação permanente na região de interesse. Entretanto, este problema pode ser contornado com uma interpolação espacial das estações mais próximas, processo também automatizado no GPTS Downloader. Por apresentar a maior consistência entre os métodos analisados, o RBF se mostrou o método mais confiável para interpolar velocidades estimadas a partir de séries temporais posicionais.

Para os 18 modelos de velocidades testados, os resultados foram significativamente piores que aqueles obtidos com as velocidades interpoladas, uma vez que os modelos utilizam dados de períodos diferentes daquele utilizado na atualização, não considerando eventos que influenciaram a posição das estações, como terremotos ou variações no comportamento das placas litosféricas.

Em relação às multivelocidades, não foi possível prever quando ou qual multivelocidade de período uniforme apresentará melhores resultados que a monovelocidade. No entanto, para séries com saltos ou mudanças de tendência, a multivelocidade com períodos distintos permite

uma melhor adequação de cada segmento ao comportamento das séries, reduzindo a discrepância após a atualização das coordenadas.

Como atividades futuras, sugere-se a implementação de um algoritmo para detectar automaticamente as mudanças de tendência nas séries e realizar a segmentação e estimativa de multivelocidades de forma automática. Ademais, considera-se relevante analisar a aplicação de modelos de PSD nas séries de estações próximas a terremotos de grande magnitude, visando corrigir os efeitos dos deslocamentos crustais não lineares, e dimensionar o impacto sobre as velocidades dessas estações. Aliado a isso, sugere-se também a avaliação das correções de deslocamento pelos sinais anuais e semianuais fornecidos pelo ITRF2020.

REFERÊNCIAS

- ABRAHAM, A.; PEDREGOSA, F.; EICKENBERG, M.; GERVAIS, P.; MUELLER, A.; KOSSAIFI, J.; GRAMFORT, A.; THIRION, B.; VAROQUAUX, G. Machine learning for neuroimaging with scikit-learn. **Frontiers in Neuroinformatics**, 8. 2014.
- AGRESTI, A. **Categorical Data Analysis**. 3. ed. Wiley, 2013.
- AKBARI, Z.; UNLAND, R. Automated Determination of the Input Parameter of DBSCAN Based on Outlier Detection. In: Iliadis, L., Maglogiannis, I. (eds) Artificial Intelligence Applications and Innovations. AIAI 2016. IFIP Advances in Information and Communication Technology, vol 475. **Springer, Cham**. 2016. DOI: https://doi.org/10.1007/978-3-319-44944-9_24.
- AKIMA, H. A new method of interpolation and smooth curve fitting based on local procedures. **Journal of the ACM (JACM)**, 17(4), 589-602. 1970.
- AKIMA, H. A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points. **ACM Transactions on Mathematical Software1 (TOMS)**, 4(2), 148-159.2. 1978
- AKKALA, A.; DEVABHAKTUNI, V.; KUMAR, A. Interpolation techniques and associated software for environmental data. **Environ. Prog. Sustain. Energy** 29 (2), 134–141, 2010.
- ALAM, B. Isolation Forest For Anomalies Detection, 2024. Disponível em: <https://hands-on.cloud/using-python-and-isolation-forest-algorithm-for-anomalies-detection/>. Acesso em: 11 jun. 2024.
- ALMEIDA, M.; OLIVEIRA, G.; DAL POZ, W. R.; MENEZES, R. R. V. Mapas de predição utilizando o interpolador krigagem a partir das componentes de velocidades das estações RBMC-IBGE. **Anais do XXVII Congresso Brasileiro de Cartografia e XXVI Expositiva** 6 a 9 de novembro de 2017, SBC, Rio de Janeiro - RJ, p. 135-139.
- ALTAMIMI, Z.; SILLARD, P.; BOUCHER, C. ITRF2000: A new release of the International Terrestrial Reference Frame for earth science applications, **J. Geophys. Res.**, 107(B10), 2214, <https://doi.org/10.1029/2001JB000561>, 2002.
- ALTAMIMI, Z.; L. MÉTIVIER; COLLILIEUX, X. ITRF2008 plate motion model, **J. Geophys. Res.**, 117(B07402), 14 pp, 2012. DOI: <https://doi.org/10.1029/2011JB008930>.
- ALTAMIMIN, Z.; REBISCHUNG, P.; MÉTLVLER, L.; COLLILLEUX, X. ITRF2014: A new release of the International Terrestrial Reference Frame modeling nonlinear station motions, **J. Geophys. Res.**, 121, 2016. 6109-6131. DOI: <https://doi.org/10.1002/2016JB013098>.
- ALTAMIMI, Z., REBISCHUNG, P., COLLILIEUX, X. et al. ITRF2020: an augmented reference frame refining the modeling of nonlinear station motions. **J Geod** 97, 47, 2023. DOI: <https://doi.org/10.1007/s00190-023-01738-w>.
- AMARAL, C. N. N.; KLEIN, I.; DE OLIVEIRA JÚNIOR, P. S. Controle de Qualidade na Remoção de Outliers para Melhoria da Estimativa do Atraso Instrumental de Receptores GNSS

.Revista Brasileira de Cartografia, [S. l.], v. 74, n. 3, p. 580–596, 2022. DOI: 10.14393/rbcv74n3-63842.

ARGUS, D.F.; GORDON, R.G. No-net-rotation model of current plate velocities incorporating plate motion model NUVEL-1, **Geophys. Res. Lett.**, 18, 2039-2042, 1991. DOI: <https://doi.org/10.1029/91GL01532>.

ARGUS, D.F.; GORDON, R.G; DEMETS, C. Geologically current motion of 56 plates relative to the no-net-rotation reference frame, **Geochemistry, Geophysics, Geosystems**, 12, No. 11, 13 pp., 2011. DOI: <https://doi.org/10.1029/2011GC003751>.

ARGUS, D.F.; GORDON, R.G; HEFLIN, M.B; MA, C.; EANES, R.J.; WILLIS, P.; PELTIER, W.R.; OWEN, S.E. The angular velocities of the plates and the velocity of the Earth's centre from space geodesy, **Geophys. J. Int.**, 18, 1-48, 2010. DOI: <https://doi.org/10.1111/j.1365-246X.2009.04463.x>.

BANVILLE, S. CSRS-PPP Version 3: Tutorial. Canadian Geodetic Survey, Surveyor General Branch, Natural Resources Canada. 2020. Disponível em: https://webapp.geod.nrcan.gc.ca/geod/tools-outils/sample_doc_filesV3/NRCan%20CSRS-PPP-v3_Tutorial%20EN.pdf. Acesso em: 03 jun. 2024.

BARATOUX, D.; *et al.* The state of planetary and space sciences in Africa, **Eos, Transactions American Geophysical Union**, v. 98. 2017. DOI: 10.1029/2017EO075833.

BARRY, D.; HARTIGAN, J. A. A Bayesian analysis for change point problems. **J. Am. Stat. Assoc.** 1993. DOI: <https://doi.org/10.2307/2290726>.

BELAY, M.; BLAKSETH, S.; RASHEED, A.; ROSSI, P..Unsupervised Anomaly Detection for IoT-Based Multivariate Time Series: Existing Solutions, Performance Analysis and Future Directions. **Sensors**. 23. 2844, 2023. DOI: 10.3390/s23052844.

BERTIGER, W. *et al.* GipsyX/RTGx, a new tool set for space geodetic operations and research. **Advances in Space Research**, v. 66, n. 3, p. 469–489, 2020. DOI: 10.1016/j.asr.2020.04.015.

BLEWITT, G.; HAMMOND, W. C.; KREEMER, C. Harnessing the GPS Data Explosion for Interdisciplinary Science, **Eos, Transactions American Geophysical Union**, 99, 2018. DOI:10.1029/2018EO104623.

BLEWITT, G.; KREEMER, C; HAMMOND, W.C.; GAZEAX, J.. MIDAS robust trend estimator for accurate GPS station velocities without step detection, **Journal of Geophysical Research**, 121, 2016, doi:10.1002/2015JB012552.

BLEWITT, G.; LAVALLÉE, D. Effect of annual signals on geodetic velocity. **Journal of Geophysical Research**. 107(B7), 2145. 2002. DOI:10.1029/2001JB000570.

BOGUSZ, J.; KŁOS, A.; GRUSZCZYŃSKA, M.; GRUSZCZYŃSKI, M. Towards Reliable Velocities of Permanent GNSS Stations. **Reports on Geodesy and Geoinformatics**. 2016. DOI: 10.1515/rgg-2016-0003.

BOGUSZ, J.; KŁOS, A.; GRZEMPOWSKI, P. *et al.* Modelling the Velocity Field in a Regular Grid in the Area of Poland on the Basis of the Velocities of European Permanent Stations. **Pure Appl. Geophys.** 171, 809–833. 2014. DOI: <https://doi.org/10.1007/s00024-013-0645-2>.

- BORRADAILE, G. J. **Statistics of Earth Science Data: Their Distribution In Time, Space, and Orientation**. Berlin: Springer, 2003.
- BOS, M. S.; FERNANDES, R. M. S.; WILLIAMS, S. D. P.; BASTOS, L. Fast error analysis of continuous GNSS observations with missing data. **Journal of Geodesy**, 87(4), 351–360, 2013. DOI:10.1007/s00190-012-0605-0.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time Series Analysis: Forecasting and Control**. 5. ed. Wiley, 2015.
- BREIMAN, L. Random Forests. *Machine Learning - Springer Nature*. 45. 5-32, 2001. DOI: 10.1023/A:1010950718922.
- BRYNS, G.; HUBERT, M.; STRUYF, A. A robust measure of skewness. **Journal of Computational and Graphical Statistics**, v. 13, n. 4, p. 996–1017, 2004.
- BURCH, T. Data collection of WGS 84 information. GPS World. 2016. Disponível em: <https://www.gpsworld.com/data-collection-of-wgs-84-information-or-is-it/>. Acesso em: 10 mar. 2024.
- BURDEN, R. L.; Faires, J. D. **Numerical Analysis**. 9. ed. Brooks/Cole, 2010.
- BUSSAB, W. O.; MORETTIN, P. A. **Estatística Básica**. 9. ed. Saraiva, 2017.
- CANNAVÒ, F.; MATTIA, M.; ROSSI, M.; PALANO, M.; BRUNO, V. A new algorithm for automatic Outlier Detection in GPS Time Series. **Geophysical Research Abstracts**. Vol. 12, EGU2010-5027, 2010.
- CASELLA, G.; BERGER, R. L. **Statistical Inference**. 2. ed. Duxbury, 2002.
- CASTRO, H. M.; FERREIRA, V. G.; FREITAS, S. R. C. Análise de deformação por variação do geopotencial: estudo de caso para o terremoto Maule (mw 8,8) com base em dados mensais da missão GRACE. **Boletim de Ciências Geodésicas**, v. 18, no 1, p.86-100, jan-mar, 2012.
- CHAPRA, S. C.; Canale, R. P. **Numerical Methods for Engineers**. 7. ed. McGraw-Hill Education, 2015.
- CHEN, C.; NOORIZADEGAN, A.; YOUNG, D.; CHEN, C. S. On the selection of a better radial basis function and its shape parameter in interpolation problems. **Applied Mathematics and Computation**. 442. 127713. 2023. DOI: 10.1016/j.amc.2022.127713.
- CONTE, S. D.; de Boor, C. **Elementary Numerical Analysis: An Algorithmic Approach**. McGraw-Hill, 1980.
- CROCETTI, L.; SCHARTNER, M.; SOJA, B. Discontinuity Detection in GNSS Station Coordinate Time Series Using Machine Learning. **Remote Sens**. 13, 3906, 2021. DOI: <https://doi.org/10.3390/rs13193906>.
- CUNHA, M. M. **Efeitos da carga oceânica no posicionamento por ponto preciso**. 2020. 186 f. Dissertação (Mestrado em Engenharia Civil) - Universidade Federal de Viçosa, Viçosa. 2020.
- DAGAR, M.; MISHRA, N.; RANI, A.; AGARWAL, S.; YADAV, J. Performance comparison of Hampel and median filters in removing deep brain stimulation artifact. **SCI**. 2018. DOI: 10.1007/978-981-10-4555-4_2.

DANEZIS C.; CHATZINIKOS, M.; KOTSAKIS, C. Linear and Nonlinear Deformation Effects in the Permanent GNSS Network of Cyprus. **Sensors**. 2020; doi:10.3390/S20061768.

DAS, T. K.; ADEPU, S.; ZHOU, J. Anomaly detection in industrial control systems using logical analysis of data. **Comput. Secur.**, 96:101935, 2020.

DAVIS, P. J. **Interpolation and approximation**. Courier Corporation, 1975.

DE BOOR, C. A Practical Guide to Spline. **Springer** Vol. 27(149), 1978. DOI: 10.2307/2006241.

DELAUNAY, B. **Sur la sphère vide**, Izvestia Akademii Nauk SSSR, Otdelenie Matematicheskikh i Estestvennykh Nauk, 7:793–800, 1934.

DEMETS, C.; R.G. GORDON; ARGUS, D.F.; STEIN, S. Effect of recent revisions to the geomagnetic reversal time scale on estimates of current plate motions, **Geophys. Res. Lett.**, 21, 2191-2194, 1994. DOI: <https://doi.org/10.1029/94GL02118>.

DEMETS, C.; GORDON, R.G.; ARGUS, D.F. Geologically current plate motions, *Geophys. J. Int.*, 181, 1-80, <https://doi.org/10.1111/j.1365-246X.2009.04491.x>, 2010; see also Erratum, **Geophys. J. Int.**, 0, 1-1, 2011. DOI: <https://doi.org/10.1111/j.1365-246X.2011.05186.x>.

DRAPER, N. R.; SMITH, H. **Applied Regression Analysis** (3rd ed.). Wiley. 1998.

DREWES, H., Combination of VLBI, SLR, and GPS determined station velocities for actual plate kinematic and crustal deformation models, *Geodynamics*, M. Feissel (Ed.), IAG Symposia, **Springer**, 1998. DOI: https://doi.org/10.1007/978-3-642-72245-5_59.

DREWES, H., The Actual Plate Kinematic and Crustal Deformation Model APKIM2005 as basis for a non-rotating ITRF, *Geodetic Reference Frames*, H. Drewes (Ed.), IAG Symposia, 134, 95-99, **Springer**, 2009. DOI: https://doi.org/10.1007/978-3-642-00860-3_15.

DREWES, H.; ANGERMANN, D. The Actual Plate Kinematic and Crustal Deformation Model 2000 (APKIM2000) as a Geodetic Reference System. **Proceedings** AIG 2001 Scientific Assembly, Budapest, 2-8 Sept 2001.

DREWES H.; HEIDBACH, O. The 2009 Horizontal Velocity Field for South America and the Caribbean. In: Kenyon S., M.C. Pacino, U. Marti (Eds.), "Geodesy for Planet Earth", **Proceedings**: IAG Symposia, 136: 657-664, 2012. DOI: 10.1007/978-3-642-20338-1_81.

DU TOIT, W. **Radial basis function interpolation**. Dissertação de mestrado, Applied Mathematics, Department of Mathematical Sciences, University of Stellenbosch, Matieland, South Africa. 2008.

EBNER, R.; FEATHERSTONE, W. E. How well can online GPS PPP post-processing services be used to establish geodetic survey control networks? **Journal of Applied Geodesy**, v. 2, n. 3, p. 149-157, 2008.

ESTER, M.; KRIEGEL, H. P.; SANDER, J.; XU, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. **Proceedings** of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, AAAI Press, pp. 226-231, 1996.

FAZILOVA, D.; EHGAMBERDIEV, SH.; KUZIN, S. Application of time series modeling to a national reference frame realization. **Geodesy and Geodynamics**, v. 9, p. 281-287, 2018. DOI: <https://doi.org/10.1016/j.geog.2018.04.003>.

FERNANDES, R.M.S., 2004. **Present-day kinematics at the Azores Gibraltar plate boundary as derived from GPS observations**, DUP, Delft, The Netherlands, ISBN 90-407-2557-8.

FERNANDES, R.M.S.; AMBROSIUS, B.A.C.; NOOMEN, R.; BASTOS, L.; WORTEL, M.; SPAKMAN, W.; GOVERS, R. The relative motion between Africa and Eurasia as derived from ITRF2000 and GPS data, **Geophys. Res. Lett.**, 30(16), 1828, 2003. DOI: 10.1029/2003GL017089.

FERNANDES, R. M. S.; MIRANDA, J. M.; MEIJNINGER, B. M. L.; BOS, M. S.; NOOMEN, R.; BASTOS, L.; AMBROSIUS, B. A. C. ; RIVA, R. E. M. Surface velocity field of the Ibero-Maghrebian segment of the Eurasia-Nubia plate boundary, **Geophysical Journal International**, Volume 169, Issue 1, Pages 315–324, 2007. DOI: <https://doi.org/10.1111/j.1365-246X.2006.03252.x>.

FERREIRA, I. O.; SANTOS, A. de P. dos; OLIVEIRA, J. C. de; MEDEIROS, N. das G.; EMILIANO, P. C. ROBUST METHODOLOGY FOR DETECTION OF SPIKES IN MULTIBEAM ECHO SOUNDER DATA. **Boletim De Ciências Geodésicas**, 25(3), e2019014. 2019. DOI: <https://doi.org/10.1590/s1982-21702019000300014>.

FREITAS, K. X. S.; NASCIMENTO, L. A.; DAL POZ, W. R. Análise da influência ionosférica e solar na acurácia do IBGE-PPP utilizando séries temporais. **Revista Brasileira de Geomática**, v. 10, p. 020, 2022.

FREITAS, K. X. S. de; DAL POZ, W. R.; NASCIMENTO, L. A. Potencial da Utilização de Multivelocidade no Processo de Atualização Temporal de Coordenadas no PPP. **Revista Brasileira de Cartografia**, [S. l.], v. 74, n. 3, p. 558–579, 2022a. DOI: 10.14393/rbcv74n3-62309.

FRITSCH, F. N.; Carlson, R. E. Monotonically piecewise cubic interpolation. **SIAM Journal on Numerical Analysis**, 17(2), 238-246, 1980.

GAGE (2025). GAGE Data Center - GPS/GNSS Data. Disponível em: <https://www.unavco.org/data/gps-gnss/gps-gnss.html#access>. Acesso em: 17 mai. 2025.

GHILANI, C. D. **Adjustment Computations - Spatial Data Analysis**. Sixth edition, John Wiley & Sons, 2017.

GILAT, A.; SUBRAMANIAM, V. **Numerical Methods for Engineers and Scientists**. 3. ed. Wiley, 2014.

GOOVAERTS, P. **Geostatistics for Natural Resources Evaluation**. New York: Oxford University Press, 1997.

GOROKHOV, O.; PETROVSKIY, M.; MASHECHKIN, I. Convolutional neural networks for unsupervised anomaly detection in text data. **Springer, Cham**, 2017.

GOYAL, R.; CHANDRA, P.; SINGH, Y. Suitability of KNN Regression in the Development of Interaction Based Software Fault Prediction Models. **Proceedings IERI**. 6. 15–21, 2014. DOI: 10.1016/j.ieri.2014.03.004.

GREENE, W. H. **Econometric Analysis**. 8. ed. Pearson, 2018.

GRIFFITHS, J.; RAY, J. Impacts of GNSS position offsets on global frame stability, **Geophysical Journal International**, Volume 204, Issue 1, Pages 480–487, 2016. DOI: <https://doi.org/10.1093/gji/ggv455>.

GRIPP, A.E.; GORDON, R.G. Current plate velocities relative to the hotspots incorporating the NUVEL-1 global plate motion model, **Geophys. Res. Lett.**, 17, 1109-1112, 1990. DOI: <https://doi.org/10.1029/GL017i008p01109>.

GRIPP, A.E.; GORDON, R.G. Young tracks of hotspots and current plate velocities, **Geophys. J. Int.**, 150, 321-361, 2002. DOI: <https://doi.org/10.1046/j.1365-246X.2002.01627.x>.

GRZECHCA, D.; TOKARZ, K.; PASZEK, K.; POLOCZEK, D. Using MEMS sensors to enhance positioning when the GPS signal disappears. **Lecture Notes in Computer Science**. Includ. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinform. 2017. DOI: https://doi.org/10.1007/978-3-319-67077-5_25.

HAMMOND, W.; BLEWITT, G.; KREEMER, C. GPS Imaging of vertical land motion in California and Nevada: Implications for Sierra Nevada uplift. **Journal of Geophysical Research: Solid Earth**. 2016. DOI: 121. 10.1002/2016JB013458.

HASSAN, A.F.; BARAKAT, S.; REZK, A. Towards a deep learning-based outlier detection approach in the context of streaming data. **J Big Data** 9, 120, 2022. DOI: <https://doi.org/10.1186/s40537-022-00670-8>.

HAWKINS, D. **Identification of Outliers**. Chapman and Hall, New York, 183p., 1980.

HE, X.; BOS, M. S.; MONTILLET, J. P.; FERNANDES, R. M. S. Investigation of the noise properties at low frequencies in long GNSS time series. **Journal of Geodesy**, v. 93, n. 9, p. 1271-1282, 2019.

HUBERT, M.; VANDERVIEREN, E. An adjusted boxplot for skewed distributions. **Computational Statistics & Data Analysis**, vol. 52, pp. 5186-5201, August 2008.

IGLEWICZ, B.; HOAGLIN, D. **How to detect and handle outliers**. Milwaukee, Wis.: ASQC Quality Press, 87p., 1993.

JAIN, P.; BAJPAI, M.; PAMULA, R. A Modified DBSCAN Algorithm for Anomaly Detection in Time-series Data with Seasonality. **The International Arab Journal of Information Technology**, 2022. DOI: 10.34028/iajit/19/1/3.

JAKOBSSON, S.; ANDERSSON, B.; EDELVIK, F. Rational radial basis function interpolation with applications to antenna design, **Journal of Computational and Applied Mathematics**, Volume 233, Issue 4, Pages 889-904, ISSN 0377-0427, 2009. DOI: <https://doi.org/10.1016/j.cam.2009.08.058>.

JESUS JUNIOR, E. J. M. de; KLEIN, I.; COSTA, R. T.; PILAPANTA, C. G. A.; OLIVEIRA JUNIOR, P. S. Estimativa e análise da aceleração horizontal em séries temporais do

posicionamento pelo GNSS em estações de monitoramento contínuo. **Revista Brasileira de Cartografia**, [S. l.], v. 76, n. 0a, 2024. DOI: 10.14393/rbcv76n0a-71296.

JI, K.; SHEN, Y. A Wavelet-Based Outlier Detection and Noise Component Analysis for GNSS Position Time Series. In: Freymueller, J.T., Sánchez, L. (eds) *Beyond 100: The Next Century in Geodesy*. International Association of Geodesy Symposia, vol 152. **Springer, Cham**. 2020. DOI: https://doi.org/10.1007/1345_2020_106.

JIANG J, DING K, LAN G. Improvement of the Estimation of the Vertical Crustal Motion Rate at GNSS Campaign Stations Based on the Information of GNSS Reference Stations. **Remote Sensing** 16(17):3144, 2024. <https://doi.org/10.3390/rs16173144>.

JOHN, H.; NAAZ, S. Credit Card Fraud Detection using Local Outlier Factor and Isolation Forest. **International Journal of Computer Sciences and Engineering**. 7, 1060-1064, 2019. DOI: 10.26438/ijcse/v7i4.10601064.

JORDAHL, K.; VAN DEN BOSSCHE, J.; FLEISCHMANN, M.; WASSERMAN, J.; MCBRIDE, J.; GERARD, J.; LEBLANC, F; et al. *geopandas/geopandas: v0.8.1 (Version v0.8.1)*. **Zenodo**, 2020. <http://doi.org/10.5281/zenodo.3946761>.

JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining Geostatistics**. Academic Press, 1978.

KAZMIER, L. J. **Estatística Aplicada à Administração e Economia**. 4ª ed, Porto Alegre: Bookman, 2007.

KHAZARADZE, G.; ECHEVERRIA, A.; ASENSIO, E. Present-day crustal deformation field of the Iberian Peninsula estimated by GPS measurements. **Física de la Tierra**; Vol 26. Retos y aplicaciones de los GNSS. 26. 35-46. 2014. DOI: 10.5209/rev_FITE.2014.v26.46970.

KIANI, M. A specifically designed machine learning algorithm for GNSS position time series prediction and its applications in outlier and anomaly detection and earthquake prediction. **Signal Processing**, 2020.

KIERULF, H.P.; STEFFEN, H.; BARLETTA, V.R.; LIDBERG, M.; JOHANSSON, J.M.; KRISTIANSEN, O.; TARASOV, L. A GNSS velocity field for geophysical applications in Fennoscandia. **Journal of Geodynamics**, 2021.

KLOPFENSTEIN JR, R. Data smoothing using a least squares fit C++ class. **ISA Transactions**, vol. 37, no. 1, pp. 3-19, 1998. DOI: [https://doi.org/10.1016/s0019-0578\(98\)00008-1](https://doi.org/10.1016/s0019-0578(98)00008-1).

KLOS, A.; BOGUSZ, J.; FIGURSKI, M.; KOSEK, W. On the Handling of Outliers in the GNSS Time Series by Means of the Noise and Probability Analysis. In: Rizos, C., Willis, P. (eds) *IAG 150 Years*. International Association of Geodesy Symposia, vol 143. **Springer, Cham**. 2015. DOI: https://doi.org/10.1007/1345_2015_78.

KLOS, A.; BOS, M.S.; BOGUSZ, J. Detecting time-varying seasonal signal in GPS position time series with different noise levels. **GPS Solut** 22, 21, 2018. DOI: <https://doi.org/10.1007/s10291-017-0686-6>.

KREEMER, C.; BLEWITT, G.; KLEIN, E.C. A geodetic plate motion and Global Strain Rate Model, **Geochemistry, Geophysics, Geosystems**, 15, 2014. DOI: 3849-3889, <https://doi.org/10.1002/2014GC005407>.

KREEMER, C.; HOLT, W.E.; HAINES, A.J. An integrated global model of present-day plate motions and plate boundary deformation, **Geophysical Journal International**, Volume 154, Issue 1, Pages 8–34, 2003. DOI: <https://doi.org/10.1046/j.1365-246X.2003.01917.x>.

LAROCHELLE, S.; CHANARD, K.; FLEITOUT, L.; FORTIN, J.; GUALANDI, A.; LONGUEVERGNE, L., et al. Understanding the geodetic signature of large aquifer systems: Example of the Ozark Plateaus in central United States. **Journal of Geophysical Research: Solid Earth**, 127, 2022.

LE THI, N.; MÄNNEL, B.; JAREMA, M.; KRISHNA SEEMALA, G.; HEKI, K.; SCHUH, H. Selection of an optimal algorithm for outlier detection in GNSS time series. **Proceedings EGU General Assembly 2021**, online, 19–30, EGU21-1598, 2021. DOI: <https://doi.org/10.5194/egusphere-egu21-1598>.

LEE, J.; PARK, J.; YOON, H. Automatic Bridge Design Parameter Extraction for Scan-to-BIM. **Applied Sciences**. 10, 2020. DOI: 7346. 10.3390/app10207346.

LEVINE, D.; M. STEPHAN, D. F.; KREHBIEL, T. C.; BERENSON, M. L. **Estatística-Teoria e Aplicações Usando Microsoft Excel** em Português. Traduzido por Teresa Cristina Padilha de Souza, 6ª ed. Rio de Janeiro: LTC, 2013.

LIMA, J. N. 2015. A utilização de filtros digitais em séries temporais GNSS. **Anais... VIII Conferência Nacional de Cartografia e Geodesia: VIII CNCG 2015**. Lisboa, Portugal, 29 e 30 de outubro de 2015.

LI, J.; MIYASHITA, K.; KATO, T. *et al.* GPS time series modeling by autoregressive moving average method: Application to the crustal deformation in central Japan. **Earth, Planets and Space** 52, 155–162, 2000. DOI: <https://doi.org/10.1186/BF03351624>.

LIU, H.; LI, L. Missing Data Imputation in GNSS Monitoring Time Series Using Temporal and Spatial Hankel Matrix Factorization. **Remote Sens.** 2022, 14, 1500. DOI: <https://doi.org/10.3390/rs14061500>.

LIU, F. T.; TING, K. M.; ZHOU, Z. Isolation Forest. **Proceedings Eighth IEEE International Conference on Data Mining, Pisa, Italy**, pp. 413-422, 2008. DOI: 10.1109/ICDM.2008.17.

LIU, F. T.; TING, K. M.; ZHOU, Z. Isolation-based anomaly detection. **ACM Transactions on Knowledge Discovery from Data (TKDD)** 6.1, 2012.

LIU, X.; JIANG, Z.; WU, Y. The Applicability of Kriging Interpolation Method in GPS Velocity Gridding and Strain Calculating. **Geomatics and Information Science of Wuhan University**, 39(4), 457-461. 2014.

LOKANAN, M. E. Financial fraud detection: the use of visualization techniques in credit card fraud and money laundering domains. **Journal of Money Laundering Control**, 2022.

LUNA, M. P.; STALLER, A.; TIERRA, A.; MOLINA, X.; TOULKERIDIS, T. Analysis of statistical interpolation methods to generate the velocities model for continental Ecuador from GNSS data. **Rev. Geogr. Venez.** 2022.

LUO, B., GAO, X.; HALABI, S. Penalized weighted proportional hazards model for robust variable selection and outlier detection. **Statistics in Medicine**, 41, 3398 – 3420, 2022.

- MAHMOUD, S.; GHAZALA, H.; RABAH, M; KHALIL1, H.; ABO-ALY, N. Evaluation of the crustal deformation in the gulf of Suez region using GPS techniques. **Proceedings of the 7th Conf. Geology of Sinai for Development Ismailia**, pp. 289-296, 2004.
- MALLINCKRODT, C. H., *et al.* . Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials, **Drug Information Journal**, 42(4), 303-319, 2008.
- MANTRI, A. Real Time Data Anomaly Detection in OTT Streaming Data Services. **European Journal of Advances in Engineering and Technology**, 8(8):98-104, 2021.
- MASSON, C.; MAZZOTTI, S.; VERNANT, P. Precision of continuous GPS velocities from statistical analysis of synthetic time series. **Solid Earth**, 2019. DOI: 10. 329-342. 10.5194/se-10-329-2019.
- MAZZOTTI, S.; DEPREZ, A.; HENRION. E.; MASSON, C.; MASSON, F.; MENUT, J.; MÉTOIS, M.; NOCQUET, J.; ROLLAND, L.; SAKIC, P.; SOCQUET, A.; SANTAMARÍA-GÓMEZ, A.; VALTY, P.; VERGNOLE, M.; VERNANT, P. Comparative analysis of synthetic GNSS time series-Bias and precision of velocity estimations. **Research Report**, RESIF. 2020.
- MICHEL, A.; SANTAMARÍA-GÓMEZ, A.; BOY, J.-P.; PEROSANZ, F.; LOYER, S. Analysis of GNSS Displacements in Europe and Their Comparison with Hydrological Loading Models. **Remote Sens.**, 13, 4523, 2021.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to Linear Regression Analysis**. 6. ed. Wiley, 2021.
- MORETTIN, P. A.; BUSSAB, W. O. **Estatística básica**. 5ª ed. São Paulo: Editora Saraiva, 526p., 2004.
- MORETTIN, P.A; TOLOI, C.M. **Análise de Séries Temporais – 2ª Edição Revista e Ampliada**. 2nd Edition, Editora Edgar Blucher, São Paulo. 2006.
- MORITZ, S.; SARDÁ, A.; BARTZ-BEIELSTEIN, T.; ZAEFFERER, M.; STORK, J. Comparison of different Methods for Univariate Time Series Imputation in R. ResearchGate, p. 1-20, out. 2015. Disponível em: <http://arxiv.org/ftp/arxiv/papers/1510/1510.03924.pdf>. Acesso em: 22 jun. 2024.
- MUNYATI, C.; SINTHUMULE, N.I. Comparative suitability of ordinary kriging and Inverse Distance Weighted interpolation for indicating intactness gradients on threatened savannah woodland and forest stands. **Environmental and Sustainability Indicators**, Volume 12, 100151, ISSN 2665-9727, 2021. DOI: <https://doi.org/10.1016/j.indic.2021.100151>.
- NAJIBI, N.; JIN, S.; WU, X. Validating the Variability of Snow Accumulation and Melting From GPS-Reflected Signals: Forward Modeling. **IEEE Transactions on Antennas and Propagation**, Vol 63, n. 6, pp. 2646-2654, 2015. doi:10.1109/TAP.2015.2414950.
- NASCIMENTO, L. A.; DAL POZ, W. R.; SANTOS, A. P.; FREITAS, K. X. S. Geração Automatizada de Séries Temporais de Dados Posicionais Utilizando o IBGE PPP. **Boletim de Ciências Geodésicas**, v. 23, p. 15 0 165, 2017.
- NASCIMENTO, L. A.; DAL POZ, W. R.; FREITAS, K. X. S. O Efeito da Carga Hidrológica e sua Influência no Deslocamento e nas Velocidades de Estações GNSS na Bacia Amazônica

Brasileira. **Revista Brasileira de Cartografia**, [S. l.], v. 73, n. 3, p. 787–806, 2021. DOI: 10.14393/rbcv73n3-57734.

NASCIMENTO, L. A.; DAL POZ, W. R.; EMILIANO, P. C.; FREITAS, K. X. S. de. Abordagem Estatística Aplicada à Investigação de Deformações Associadas ao Efeito de Carga Hidrológica na Amazônia Brasileira. **Revista Brasileira de Cartografia**, [S. l.], v. 73, n. 4, p. 1017–1035, 2021. DOI: 10.14393/rbcv73n4-60356.

NETO, T. F. C.; CASCUDO, O. Imputação de dados ausentes em séries temporais de carbonatação dos concretos. **Ambiente Construído**, v. 24, e136386, Porto Alegre, jan./dez. 2024. ISSN 1678-8621 Associação Nacional de Tecnologia do Ambiente Construído. DOI: <http://dx.doi.org/10.1590/s1678-86212024000100748>.

NEWSOM, J. T. Multiple Regression and Multivariate Quantitative Methods. Portland State University. Disponível em: https://web.pdx.edu/~newsomj/mvclass/ho_missing.pdf. Acesso em: 20 maio. 2025.

NGL (2025) – NEVADA GEODETIC LABORATORY. Disponível em: <https://geodesy.unr.edu/>. Acesso em: 14 out 2025.

NGUYEN, H.D.; TRAN, T.D. Detecting outliers in GNSS position time series using machine learning techniques (in Vietnamese). **Journal of Mining and Earth Sciences**. 64, 4 (Aug, 2023), 22-30. 2023. DOI: [https://doi.org/10.46326/JMES.2023.64\(4\).03](https://doi.org/10.46326/JMES.2023.64(4).03).

NISTOR, S.; SUBA, N.-S.; MACIUK, K.; KUDRYS, J.; NASTASE, E.I.; MUNTEAN, A. Analysis of Noise and Velocity in GNSS EPN-Repro 2 Time Series. **Remote Sens**, 13, 2783. 2021. DOI: <https://doi.org/10.3390/rs13142783>.

NV5 (2024) - NV5 GEOSPATIAL SOFTWARE – Griddata. Disponível em: <https://www.nv5geospatialsoftware.com/docs/GRIDDATA.html>. Acesso em: 02 mai. 2024.

PAHLEVI, A. M.; PRIJATNA, K.; MEILANO, I.; DAN IBNU, S. Investigation of the Solid Earth Tide Based on GPS Observation and Superconducting Gravimeter Data. **Jurnal Ilmiah Geomatika**. 2017; doi:10.24895/JIG.2016.22-1.488.

PANDAS (2025a) – Pandas Pydata – DataFrame interpolate. Disponível em: <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.interpolate.html>. Acesso em: 13 mai. 2024.

PANDAS (2025b) – Pandas Pydata - Rolling. Disponível em: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.rolling.html>. Acesso em: 16 mai. 2024.

PANDAS (2025c) – Pandas Pydata - Mean. Disponível em: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.mean.html>. Acesso em: 16 mai. 2024.

PARK, C.H. A Comparative Study for Outlier Detection Methods in High Dimensional Text Data. **Journal of Artificial Intelligence and Soft Computing Research**, 13, 5 – 17, 2022.

PARK J.; GREJNER-BRZEZINSKA, D. A.; VON FRESE, R. R. B. GPS Discrimination of Traveling Ionospheric Disturbances from Underground Nuclear Explosions and

Earthquakes. **Journal of the Institute Navigation**, 61, p. 125-134. 2014; DOI:10.1002/NAVI.56.

PAPASTEFANOPOULOS, V.; LINARDATOS, P.; KOTSIANTIS, S. Unsupervised Outlier Detection: A Meta-Learning Algorithm Based on Feature Selection. **Electronics** 10, 2236. 2021. DOI: <https://doi.org/10.3390/electronics10182236>.

PAVANI, A.; TEJASRI, D.; REVATHI, R.; RAMESH, K.; KOTESWARA RAO, S.; Lakshminarayana, S. Application of Modified Periodogram Algorithm on GPS TEC for Earthquakes. **Indian Journal of Science and Technology**. 9. 2016. DOI: 10.17485/ijst/2016/v9i17/93117.

PEDREGOSA *et al.* Scikit-learn: Machine Learning in Python. **JMLR** 12, pp. 2825-2830, 2011.

PEHLIVAN, H. A novel outlier detection method based on Bayesian change point analysis and Hampel identifier for GNSS coordinate time series. **EURASIP Journal on Advances in Signal Processing**. 2024, 44. 2024. DOI: <https://doi.org/10.1186/s13634-023-01097-w>.

PEREIRA, F.M.; SOFIA, R.C. An Analysis of ML-Based Outlier Detection from Mobile Phone Trajectories. **Future Internet** 2023, 15, 4, 2022. DOI: <https://doi.org/10.3390/fi15010004>.

PESCHON, M. DBSCAN & Outlier Detection. Rpubs R Studio. 2020. Disponível em: <https://rpubs.com/gauxgaux/580701>. Acesso em: 12 jun. 2024.

PETIT, G; LUZUM, B (eds.). International Earth Rotation and Reference Systems Service (IERS). IERS Technical. Note No. 36. IERS Conventions. Frankfurt am Main, 2010. Disponível em: https://www.iers.org/SharedDocs/Publikationen/EN/IERS/Publications/tn/TechnNote36/tn36.pdf?__blob=publicationFile&v=1. Acesso em: 28 mar. 2024.

PRAWIRODIRDJO, L.; BOCK, Y. Instantaneous global plate motion model from 12 years of continuous GPS observations, **J. Geophys. Res.**, 109, B08405, 2004. DOI: 10.1029/2003JB002944, 2004.

PRESS, W. H., *et al.* **Numerical recipes: The art of scientific computing**. 3. Ed. Cambridge university press, 2007.

PROL, F. S.; JUNIOR, J. M.; NIEVINSKI, F. G.; GOMES, R. L. E FILHO, A. C. P. Transformação entre referenciais e cálculo de velocidades através do aplicativo Web Trelvel. **Revista Brasileira de Cartografia**. v. 66, n. 3, p. 569-579, 2014.

QIANQIAN, Z.; QINGMING, G.. Bayesian methods for outliers detection in GNSS time series. **Journal of Geodesy**, 87, 2013. DOI: 10.1007/s00190-013-0640-5.

RAMOS, M. DE P.; DAL POZ, W. R.; CARVALHO, A. S. Determinação de Velocidades das Estações da RBMC com Uso do Software SARI. **Revista Brasileira de Cartografia**, v. 73, n. 2, p. 453-469, 6 abr. 2021.

RAMOS, M. DE P.; DAL POZ, W. R.; CARVALHO, A. S. Propagação de Incertezas no Processo de Compatibilização de Referenciais e Época de Coordenadas GNSS. **Revista Brasileira de Cartografia**, [S. l.], v. 74, n. 2, p. 305–321, 2022. DOI: 10.14393/rbcv74n2-64767.

RAMASWAMY, S.; RASTOGI, R.; SHIM, K. Efficient algorithms for mining outliers from large data sets. **Proceedings** of the 2000 ACM SIGMOD International Conference on Management of Data, Dallas, TX, USA; pp. 427–438, 2000.

ROBERTS, E.A.; SHELEY, R.L.; LAWRENCE, R.L. Using sampling and inverse distance weighted modeling for mapping invasive plants. **West. N. Am. Naturalist**. 64, 312–323, 2004.

RODRIGUEZ-GALIANO, V.F.; CHICA-OLMO, M.; CHICA-RIVAS, M. Predictive modelling of gold potential with the integration of multisource information based on random forest: A case study on the Rodalquilar area, Southern Spain. **International Journal of Geographical Information Science**, 28, 1336–1354, 2014.

RODRIGUEZ-GALIANO, V.; SANCHEZ-CASTILLO, M.; CHICA-OLMO, M.; CHICA-RIVAS, M. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. **Ore Geology Reviews**, 71, 804–818, 2015.

ROMANO, J.P.; WOLF, M. Resurrecting weighted least squares, **Journal of Econometrics**, Volume 197, Issue 1, Pages 1-19, ISSN 0304-4076, 2017. DOI: <https://doi.org/10.1016/j.jeconom.2016.10.003>.

SAKIC, P., MANSUR, G., CHAIYAPORN, K., and BALLU, V. The geodeZXYX toolbox: a versatile Python 3 toolbox for geodetic-oriented purposes. V. 4.0. **GFZ Data Services**, 2019. DOI: <http://doi.org/10.5880/GFZ.1.1.2019.002>.

SALEH, J., BENNETT, R.A. & WILLIAMS, S.D.P. Minimum-entropy velocity estimation from GPS position time series. **J Geod** 98, 11, 2024.

SÁNCHEZ L.; DREWES H. VEMOS2015: Velocity and deformation model for Latin America and the Caribbean, 2016. DOI: 10.1594/PANGAEA.863131. supplement to: SÁNCHEZ L. DREWES H. Crustal deformation and surface kinematics after the 2010 earthquakes. Latin America. **Journal of Geodynamics**. 2016. DOI: 10.1016/j.jog.2016.06.005.

SANTAMARÍA-GÓMEZ, A. SARI: interactive GNSS position time series analysis software. **GPS Solutions**, v. 23, n. 2, p. 52, 2019. DOI. 10.1007/s10291-019-0846-y.

SAPUCCI, L. F.; MONICO, J. F. G. Transformação de Helmert generalizada no posicionamento de alta precisão: fundamentação teórica e exemplificações. **Revista Brasileira de Geofísica**. Presidente Prudente-SP, v.18, n.2, pp.161-172, 2001. DOI: <https://doi.org/10.1590/S0102-261X2000000200005>.

SAVCHYN, I.; BRUSAK, I.; TRETAK, K. Analysis of recent Antarctic plate kinematics based on GNSS data. **Geodesy and Geodynamics**. 14. 2023. DOI: 10.1016/j.geog.2022.08.004.

SCHUBERT, E.; SANDER, J.; ESTER, M.; KRIEGEL, H. P.; XU, X. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. **ACM Transactions on Database Systems (TODS)**, 42(3), 19, 2017.

SCIKIT-LEARN (2025a) - Isolation Forest. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html>. Acesso em: 03 abr. 2025.

SCIKIT-LEARN (2025b) - DBSCAN. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html>. Acesso em: 03 abr. 2025.

SEABOLD, S.; PERKTOLD, J. *Statsmodels: Econometric and Statistical Modeling with Python*. SciPy, 2010.

SELLA, G.F.; DIXON, T.H.; MAO, A. REVEL: A model for recent plate velocities from space geodesy, **J. Geophys. Res.**, 107, B4, 2002. DOI: <https://doi.org/10.1029/2000JB000033>.

SEN, P. K. Estimates of the regression coefficient based on Kendall's tau. **Journal of the American Statistical Association**. 63, 1379–1389. 1968.

SHEN, Y.; YANG, L. The Importance of Suppressing Complete Reconstruction in Autoencoders for Unsupervised Outlier Detection. **ArXiv**, abs/2211.03054, 2022.

SHETTA, O.; NIRANJAN, M. Robust subspace methods for outlier detection in genomic data circumvents the curse of dimensionality. **Royal Society Open Science**, 7(2):190714, 2020.

SHUMWAY, R.; STOFFER, D. **Time Series Analysis and Its Applications** With R Examples. 3th edition. 2011. DOI: 10.1007/978-1-4419-7865-3.

SIRGAS (2025a). Sistema de Referência Geocêntrico para as Américas, Modelo de velocidade para o SIRGAS. Disponível em: <https://www.sirgas.org/pt/velocity-model/>. Acesso em: 18 mai. 2025.

SIRGAS (2025b). Sistema de Referência Geocêntrico para as Américas, Sistema de Referência Geocêntrico. Disponível em: <https://www.sirgas.org/pt/velocity-model/using-velocities-gnss/>. Acesso em: 19 mai. 2025.

SONG, Y.; LIANG, J.; LU, J.; ZHAO, X. An efficient instance selection algorithm for k nearest neighbor regression. **Neurocomputing**, 251, 2017. DOI: 10.1016/j.neucom.2017.04.018.

SUN, T.; LI, H.; WU, K.; CHEN, F.; ZHU, Z.; HU, Z. Data-Driven Predictive Modelling of Mineral Prospectivity Using Machine Learning and Deep Learning Methods: A Case Study from Southern Jiangxi Province, China. **Minerals** 10, 102, 2020. DOI: <https://doi.org/10.3390/min10020102>.

STATSMODELS (2025) - statsmodels.tsa.seasonal.seasonal_decompose. Disponível em: https://www.statsmodels.org/dev/generated/statsmodels.tsa.seasonal.seasonal_decompose.html. Acesso em: 26 mar. 2025.

STOER, J.; BULIRSCH, R. **Introduction to Numerical Analysis** 3. ed. Springer, 2002.

QIANQIAN, Z.; QINGMING, G. Bayesian methods for outliers detection in GNSS time series. **J Geod** 87, 609–627, 2013. DOI: <https://doi.org/10.1007/s00190-013-0640-5>.

TAPIQUÉN, C. E. P. "World Countries.shp". Based on shapes from Environmental Systems Research Institute (ESRI). Orogénesis Soluciones Geográficas. Porlamar, Venezuela 2015. Disponível em: <http://tapiquen-sig.jimdo.com>. Acesso em: 03 fev. 2025.

THEIL, H. A rank-invariant method of linear and polynomial regression analysis, **Indagationes Mathematicae**. 12, 85–91, 1950.

THOMAS, T.; RAJABI, E. A systematic review of machine learning-based missing value imputation techniques. **Data Technologies and Applications**, v. 55, n. 4, 2021.

TORGE, W. **Geodesy**. 3ª edição. Berlim-New York: Walter de Gruyter, 2001.

TRAN, D.T., **Analyse Rapide et Robust des Solutions GPS Pour la Technique**, PhD Thesis, National Library, 2013.

TRAN, T.; DAO, T.; VU, T.; LUONG, D.; VU, C.; BUI, S.; HA, H. Outlier detection in GNSS position time series. **Science and Technology Development Journal**, 19(2), 43-50. 2016. DOI: <https://doi.org/https://doi.org/10.32508/stdj.v19i2.665>.

UNAVCO (2025). Plate Motion Calculator. Disponível em: <http://www.unavco.org/software/geodetic-utilities/plate-motion-calculator/plate-motion-calculator.html> . Acesso em: 15 mai. 2025.

VANDERVIJREN, E.; HUBERT, M. An adjusted boxplot for skewed distributions. **Proceedings in Computational Statistics**, p. 1933-1940, 2004.

WANG, G. The 95% Confidence Interval for GNSS-Derived Site Velocities. **Journal of Surveying Engineering**. 148, 2021. DOI: 10.1061/(ASCE)SU.1943-5428.0000390.

WATSON, D.F.; PHILIP, G.M., A refinement of inverse distance weighted interpolation. **Geo Process**. 2, 315–327, 1985.

WITMAN, S. Understanding tectonic processes following great earthquakes. **Eos, Transactions American Geophysical Union**, volume 97. 2016. DOI: 10.1029/2016EO063695.

WOOLDRIDGE, J. M. **Introductory Econometrics: A Modern Approach**. 6. ed. Cengage Learning, 2016.

WOOLDRIDGE, J. M. **Introductory Econometrics: A Modern Approach**. 7. ed. Cengage Learning, 2019.

YAMAMOTO, J. K.; LANDIM, P. M. B. **Geoestatística: Conceitos e Aplicações**. Oficina de Textos, 2013

ZENG, Y. GPS Velocity Field of the Western United States for the 2023 National Seismic Hazard Model Update. **Seismological Research Letters**, 93 (6): 3121–3134, 2022. DOI: <https://doi.org/10.1785/0220220180>.

ZHAO, Y.; NASRULLAH, Z.; LI, Z. PyOD: A Python Toolbox for Scalable Outlier Detection. **Journal of machine learning research (JMLR)**, 20(96), pp.1-7, 2019.

ZHANG, N.; ZHOU, K.; Li, D. Back-propagation neural network and support vector machines for gold mineral prospectivity mapping in the Hatu region, Xinjiang, China. **Earth. Sci. Inform**, 11, 553–566, 2018.

APÊNDICES

APÊNDICE A – Informações sobre as estações utilizadas.

| Estação | Latitude (°) | Longitude (°) | Altitude (m) | Estado | Placa Litoférica |
|----------------|---------------------|----------------------|---------------------|---------------|-------------------------|
| AGMT | 34,5942819 | -116,4293773 | 1337,8313 | California | Norte-Americana |
| AVRY | 34,4683010 | -117,1539700 | 888,9007 | California | Norte-Americana |
| AZRY | 33,5400716 | -116,6296946 | 1265,6972 | California | Pacífico |
| AZU1 | 34,1260177 | -117,8964841 | 144,7630 | California | Pacífico |
| BARH | 44,3950478 | -68,2216892 | 6,7190 | Maine | Norte-Americana |
| BBRY | 34,2642784 | -116,8842399 | 2051,0533 | California | Norte-Americana |
| BGIS | 33,9671167 | -118,1596934 | 2,8263 | California | Pacífico |
| BILL | 33,5782421 | -117,0645971 | 470,0478 | California | Pacífico |
| BKMS | 33,9622567 | -118,0946950 | 10,9928 | California | Pacífico |
| BLYT | 33,6104155 | -114,7148499 | 85,9164 | California | Norte-Americana |
| BMHL | 34,2514455 | -116,0529771 | 722,6285 | California | Norte-Americana |
| BRAN | 34,1848934 | -118,2770469 | 246,2452 | California | Pacífico |
| BSRY | 34,9186122 | -117,0119914 | 613,4944 | California | Norte-Americana |
| BURN | 42,7795020 | -117,8435283 | 1180,9163 | Oregon | Norte-Americana |
| CAND | 35,9393525 | -120,4336951 | 568,3926 | California | Norte-Americana |
| CAST | 39,1910226 | -110,6773123 | 2245,0478 | Utah | Norte-Americana |
| CBHS | 34,1385599 | -118,6298019 | 284,6048 | California | Pacífico |
| CHIL | 34,3334219 | -118,0259963 | 1567,5030 | California | Pacífico |
| CHMS | 34,6404617 | -117,8276994 | 760,3164 | California | Norte-Americana |
| CIT1 | 34,1367077 | -118,1272822 | 215,3390 | California | Pacífico |
| CJMS | 34,3137986 | -117,4793806 | 933,3437 | California | Pacífico |
| CJTR | 34,8216951 | -92,2730360 | 126,0010 | Arkansas | Norte-Americana |
| CLAR | 34,1099273 | -117,7088068 | 373,6278 | California | Pacífico |
| COON | 40,6525920 | -112,1210100 | 1610,0151 | Utah | Norte-Americana |
| CORV | 44,5855356 | -123,3046037 | 105,6137 | Oregon | Norte-Americana |
| CRFP | 34,0390513 | -117,0996801 | 688,8054 | California | Pacífico |
| CRHS | 33,8235027 | -118,2727626 | -23,5406 | California | Pacífico |
| CSDH | 33,8614762 | -118,2567137 | -9,1761 | California | Pacífico |
| CTMS | 34,1241006 | -116,3704246 | 966,4649 | California | Norte-Americana |
| CVHS | 34,0820101 | -117,9017141 | 119,1062 | California | Pacífico |
| DHLG | 33,3898061 | -115,7880267 | -83,0537 | California | Pacífico |
| DSSC | 33,7333304 | -116,7120872 | 1660,8543 | California | Pacífico |
| ECHO | 37,9155344 | -114,2642465 | 1684,1159 | Nevada | Norte-Americana |
| EGAN | 39,3452443 | -114,9388534 | 1998,2799 | Nevada | Norte-Americana |
| ELSC | 34,0297326 | -118,2084293 | 61,1979 | California | Pacífico |
| EOUT | 41,2531972 | -111,9289075 | 1677,8469 | Utah | Norte-Americana |
| EWPP | 34,1041973 | -117,5255821 | 330,4771 | California | Pacífico |
| FERN | 35,3418726 | -112,4547360 | 1767,9154 | Arizona | Norte-Americana |
| FZHS | 34,8001927 | -118,8933778 | 1273,7076 | California | Pacífico |
| GARL | 40,4165271 | -119,3554556 | 1640,2405 | Nevada | Norte-Americana |

| | | | | | |
|------|------------|--------------|-----------|------------|-----------------|
| GMRC | 34,7839982 | -115,6602281 | 1298,2089 | California | Norte-Americana |
| GOSH | 40,6401693 | -114,1796965 | 1560,1947 | Nevada | Norte-Americana |
| HCMN | 34,7547742 | -116,4300734 | 568,7487 | California | Norte-Americana |
| HNPT | 38,5888164 | -76,1303728 | -27,9520 | Maryland | Norte-Americana |
| IMPS | 34,1575639 | -115,1450987 | 563,2562 | California | Norte-Americana |
| JPLM | 34,2048193 | -118,1732239 | 423,9860 | California | Pacífico |
| KNOL | 37,6591179 | -118,9791717 | 2651,7863 | California | Norte-Americana |
| KOKB | 22,1262612 | -159,6649249 | 1167,3489 | Hawaii | Pacífico |
| LAND | 35,8997888 | -120,4732802 | 568,5474 | California | Pacífico |
| LAPC | 34,1819206 | -118,5746473 | 207,8248 | California | Pacífico |
| LBC1 | 33,8320684 | -118,1371800 | -21,9197 | California | Pacífico |
| LBC2 | 33,7916078 | -118,1732386 | -28,4750 | California | Pacífico |
| LDES | 34,2673394 | -116,4328023 | 978,0306 | California | Norte-Americana |
| LEWI | 40,4035133 | -116,8618718 | 2916,7179 | Nevada | Norte-Americana |
| LFRS | 34,0950692 | -118,4128221 | 146,9010 | California | Pacífico |
| LINC | 37,6371928 | -119,0172858 | 3057,0909 | California | Norte-Americana |
| LINJ | 34,6620213 | -118,1391985 | 721,2434 | California | Norte-Americana |
| LLAS | 34,5859746 | -117,8379292 | 812,4888 | California | Norte-Americana |
| LMUT | 40,2614208 | -111,9283130 | 2311,9890 | Utah | Norte-Americana |
| LORS | 34,1333261 | -117,7540669 | 448,8940 | California | Pacífico |
| LVMS | 34,7336049 | -119,1036829 | 1543,0767 | California | Pacífico |
| MAUI | 20,7066529 | -156,2570211 | 3062,1081 | Hawaii | Pacífico |
| MAWY | 44,9734283 | -110,6893001 | 1824,2526 | Wyoming | Norte-Americana |
| MERC | 36,6325126 | -115,9790149 | 1256,9266 | Nevada | Norte-Americana |
| MIDA | 35,9219126 | -120,4588287 | 570,6861 | California | Pacífico |
| MKEA | 19,8013522 | -155,4563371 | 3754,6757 | Hawaii | Pacífico |
| MLFP | 33,9184025 | -117,3179552 | 472,9683 | California | Pacífico |
| MPWD | 34,2955100 | -118,8780130 | 190,2413 | California | Pacífico |
| MUSB | 37,1699411 | -119,3093488 | 2042,5177 | California | Norte-Americana |
| MWTP | 37,6405194 | -118,9447284 | 2317,4112 | California | Norte-Americana |
| NAIU | 41,0156730 | -112,2296253 | 1455,7861 | Utah | Norte-Americana |
| NEWS | 39,6856172 | -117,5089348 | 1770,2665 | Nevada | Norte-Americana |
| NLIB | 41,7715910 | -91,5748943 | 207,0457 | Iowa | Norte-Americana |
| OAES | 34,1410055 | -116,0677352 | 604,7710 | California | Norte-Americana |
| OPBL | 34,3698829 | -115,9180561 | 1226,5397 | California | Norte-Americana |
| OPCP | 34,3671297 | -116,0833713 | 1096,6545 | California | Norte-Americana |
| OPCX | 34,4300827 | -116,1494830 | 1133,8124 | California | Norte-Americana |
| OPRD | 34,5330318 | -116,2922778 | 1400,2095 | California | Norte-Americana |
| ORES | 34,7390581 | -120,2785549 | 143,8692 | California | Pacífico |
| OVLS | 34,3273582 | -119,1419435 | 69,2122 | California | Pacífico |
| PMTN | 37,8313721 | -119,0569265 | 2401,6123 | California | Norte-Americana |
| POMM | 35,9199094 | -120,4784318 | 597,1004 | California | Pacífico |
| PPBF | 33,8357256 | -117,1820859 | 428,0938 | California | Pacífico |
| PSDM | 34,0917640 | -117,8070881 | 278,3796 | California | Pacífico |
| PTGV | 36,4134713 | -89,6996823 | 58,0318 | Missouri | Norte-Americana |
| PVHS | 33,7794525 | -118,3721610 | 259,5925 | California | Pacífico |

| | | | | | |
|------|------------|--------------|-----------|------------|-----------------|
| RBUT | 40,7810635 | -111,8087597 | 1667,7619 | Utah | Norte-Americana |
| RDOM | 37,6770683 | -118,8979411 | 2417,4226 | California | Norte-Americana |
| REDM | 44,2597678 | -121,1478801 | 919,8552 | Oregon | Norte-Americana |
| RHCL | 34,0190497 | -118,0261668 | 176,8788 | California | Pacífico |
| RTHS | 34,0891495 | -117,3533313 | 328,6826 | California | Pacífico |
| RUBY | 40,6172108 | -115,1228332 | 1881,7449 | Nevada | Norte-Americana |
| S300 | 37,6665075 | -121,5582661 | 495,7423 | California | Norte-Americana |
| SEDR | 48,5215584 | -122,2238470 | 29,8645 | Washington | Norte-Americana |
| SHIN | 40,5916792 | -120,2250420 | 1377,3276 | California | Norte-Americana |
| SHLD | 41,8683678 | -119,0156833 | 1457,5588 | Nevada | Norte-Americana |
| SHOS | 35,9713460 | -116,2989835 | 582,4042 | California | Norte-Americana |
| SLMS | 33,2922271 | -115,9778396 | -45,9438 | California | Pacífico |
| SMEL | 39,4256442 | -112,8449266 | 1422,1506 | Utah | Norte-Americana |
| SPMS | 33,9926531 | -117,8487719 | 207,0384 | California | Pacífico |
| TABL | 34,3818482 | -117,6782795 | 2228,0215 | California | Pacífico |
| TIVA | 36,9348498 | -116,2297047 | 2034,1744 | Nevada | Norte-Americana |
| TONO | 38,0971945 | -117,1840394 | 2066,0265 | Nevada | Norte-Americana |
| TORP | 33,7977967 | -118,3306001 | -5,2137 | California | Pacífico |
| TOST | 34,2479544 | -118,8366611 | 274,3853 | California | Pacífico |
| USGC | 33,0300625 | -116,0853313 | 133,4317 | California | Pacífico |
| VNCX | 34,2931919 | -118,4845366 | 328,5780 | California | Pacífico |
| VNPS | 34,5014803 | -118,1211552 | 961,8968 | California | Pacífico |
| WCHS | 34,0618967 | -117,9111073 | 100,1079 | California | Pacífico |
| WHC1 | 33,9798838 | -118,0311659 | 94,2844 | California | Pacífico |
| WLSN | 34,2261203 | -118,0559093 | 1705,2480 | California | Pacífico |
| WMAP | 34,2594230 | -118,4142881 | 268,1044 | California | Pacífico |
| WRHS | 33,9581506 | -118,4276035 | 7,8601 | California | Pacífico |

APÊNDICE B – Multivelocidades de períodos não uniformes

| Estação | Segmento 1 | Segmento 2 | Segmento 3 | Segmento 4 |
|----------------|----------------------------|----------------------------|----------------------------|----------------------------|
| BMHL | 01/01/2000 - 31/12/2003 | 01/01/2004 - 30/04/2010 | 01/05/2010 - 31/12/2014 | 01/01/2015 - 31/12/2019 |
| DHLG | 01/01/2000 - 31/12/2003 | 01/01/2004 - 31/12/2011 | 01/01/2012 - 31/12/2019 | |
| MAWY | 01/01/2000 - 31/12/2003 | 01/01/2004 - 31/08/2010 | 01/09/2010 - 31/12/2013 | 01/01/2014 - 31/12/2019 |
| NAIU | 01/01/2000 - 31/12/2003 | 01/01/2004 - 31/12/2012 | 01/01/2013 - 31/12/2019 | |
| NLIB | 01/01/2000 - 31/12/2003 | 01/01/2004 - 31/12/2011 | 01/01/2012 - 31/12/2019 | |
| OPBL | 01/01/2000 - 31/12/2003 | 01/01/2004 - 31/12/2010 | 01/01/2011 - 31/12/2019 | |
| OPCP | 01/01/2000 - 30/06/2004 | 01/07/2004 - 28/02/2014 | 01/03/2014 - 31/12/2019 | |
| OPCX | 01/01/2000 - 30/06/2003 | 01/07/2003 - 31/12/2009 | 01/01/2010 - 31/12/2019 | |
| OPRD | 01/01/2000 - 31/12/2003 | 01/01/2004 - 30/04/2010 | 01/05/2010 - 31/05/2014 | 01/06/2014 - 31/12/2019 |
| SLMS | 01/01/2000 - 30/06/2004 | 01/07/2004 - 31/12/2010 | 01/01/2011 - 31/12/2019 | |