



## O Efeito das Covariâncias entre os Componentes de Linha Base sobre a Confiabilidade de Redes GNSS: Resultados para uma Rede com Alta Redundância

### *The Effect of Covariance between Baseline Components on the Reliability of GNSS Networks: Results for a Highly Redundancy Network*

Maria Luísa Silva Bonimani<sup>1</sup>, Vinicius Francisco Rofatto<sup>2</sup>, Marcelo Tomio Matsuoka<sup>1,2,3</sup>, Ivandro Klein<sup>4,5</sup>, Maurício Roberto Veronez<sup>6</sup> e Luiz Gonzaga da Silveira Júnior<sup>6</sup>

<sup>1</sup> Universidade Federal de Uberlândia, Programa de Pós-Graduação em Agricultura e Informações Geoespaciais, Monte Carmelo, Brasil, [malubonimani@hotmail.com](mailto:malubonimani@hotmail.com), [tomiomatsuoka@gmail.com](mailto:tomiomatsuoka@gmail.com)

ORCID: <https://orcid.org/0000-0002-3314-3230>

ORCID: <https://orcid.org/0000-0002-2630-522X>

<sup>2</sup> Universidade Federal de Uberlândia, Instituto de Geografia, Monte Carmelo, Brasil, [vprofatto@gmail.com](mailto:vprofatto@gmail.com)

ORCID: <https://orcid.org/0000-0003-1453-7530>

<sup>3</sup> Universidade Federal de Uberlândia, Programa de Pós-Graduação em Geografia, Uberlândia, Brasil

<sup>4</sup> Instituto Federal de Santa Catarina, Departamento Acadêmico de Construção Civil, Florianópolis, Brasil, [ivandroklein@gmail.com](mailto:ivandroklein@gmail.com)

ORCID: <https://orcid.org/0000-0003-4296-592X>

<sup>5</sup> Universidade Federal do Paraná, Programa de Pós-Graduação em Ciências Geodésicas, Curitiba, Brasil

<sup>6</sup> Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, São Leopoldo, Brasil,

[veronez@unisinos.br](mailto:veronez@unisinos.br), [lgonzagajr@gmail.com](mailto:lgonzagajr@gmail.com)

ORCID: <https://orcid.org/0000-0002-5914-3546>

ORCID: <https://orcid.org/0000-0002-7661-2447>

Recebido: 11.2020 | Aceito: 12.2020

**Resumo:** A mais recente versão da teoria da confiabilidade proposta por Rofatto (2020) tem sido utilizada para descrever a capacidade de um sistema de medição em detectar, identificar e remover *outliers* a um certo nível de probabilidade. Entretanto, a teoria tem sido aplicada em redes nivelamento. Aqui, por outro lado, aplicamos a teoria no contexto de redes baseadas nos sistemas de posicionamento por satélites GNSS (*Global Navigation Satellite System*), a partir de dados reais coletados em campo. Testamos se as covariâncias entre as componentes da linha base têm efeito sobre a confiabilidade. Verificamos que as covariâncias entre as componentes da linha base aumentam a taxa de sucesso na identificação de *outlier* e, portanto, aumentam a confiabilidade da rede. O menor *outlier* identificável – ao nível de 80% de correta identificação – teve uma redução média de ~30% para as componentes  $\Delta X$  e  $\Delta Y$ , e ~14% para  $\Delta Z$  em comparação ao cenário com covariâncias nulas. O aumento do nível de significância melhora a confiabilidade em ambos os cenários (covariâncias nulas e não-nulas) na mesma proporção. Porém, para altos níveis de significância ( $\alpha > 0,1$ ) e sistemas com boa redundância ( $ri > 0,5$ ), a confiabilidade para um modelo estocástico com covariâncias nulas se aproxima do caso em que as covariâncias não são nulas. Na ausência de um modelo estocástico mais realista (covariâncias não-nulas) e para sistemas com boa redundância local ( $ri > 0,5$ ), deve-se optar por regiões críticas maiores ( $\hat{k} < 2,8$ ).

**Palavras-chave:** Confiabilidade. *Outlier*. Monte Carlo. Controle de Qualidade. Rede. Posicionamento por Satélites.

**Abstract:** The most recent version of the reliability theory proposed by Rofatto (2020) has been applied to describe the ability of a measurement system to detect, identify and remove outliers for a given probability level. However, the applications of that theory have been addressed to the simulated leveling networks. Here, however, we apply the theory within the context of satellite positioning-based network with real data. We have tested whether the covariances between the baseline components have effect or not on the reliability. We show that covariances increase the success rate in terms of outlier identification and, therefore, improves the reliability of the network. In comparison to a null covariance scenario, the minimal identifiable outlier for 80% of correct identification has decreased about ~30% for the  $\Delta X$  and  $\Delta Y$  components, and ~14% for  $\Delta Z$  component. The increase of the significance level had the same proportion of improvement in both scenarios (null and non-null covariance). For high significance levels ( $\alpha > 0,1$ ) and measurements systems with good redundancy ( $ri > 0,5$ ), however, the reliability of a stochastic model with null covariances is actually close to that of a full covariances scenario. In the absence of a more realistic stochastic model (e.g.: full covariance between baseline components) and systems with good local redundancy ( $ri > 0,5$ ), one should choose high critical values ( $\hat{k} < 2,8$ ).

**Keywords:** Reliability. *Outlier*. Monte Carlo. Quality control. Network. Satellite Positioning.

## 1 INTRODUÇÃO

*Outlier* é uma observação que se afastou de seu valor mais provável a ponto de não pertencer ao modelo matemático (funcional e estocástico) estipulado (ROFATTO, 2020). Portanto, os *outliers* devem ser tratados de forma adequada a fim de garantir a confiabilidade dos resultados.

A mais de meio século, o procedimento estatístico *Data-Snooping* (DS) – proposto por Baarda (1968) – vem sendo utilizado na Geodésia como uma prática padrão para remoção de possíveis *outliers* nos dados observados (ROFATTO et al., 2020a). Ao fazer cálculos de ajustamento de observações geodésicas – baseados no Método dos Mínimos Quadrados (MMQ) –, este procedimento é um dos mais utilizados para identificar possíveis *outliers* (LEHMANN; VOß-BÖHME, 2017). Assim, qualquer software adequado de processamento de dados geodésicos – que opera sob o critério dos mínimos quadrados – tem implementado este procedimento. Importante notar que grande parte dos livros didáticos de Geodésia tem um capítulo sobre o procedimento *Data-Snooping*, por exemplo: Koch (1999), Teunissen (2006), Monico (2008), Ghilani (2010), Leick, Rapoport e Tatarnikov (2015), entre outros.

O DS é frequentemente aplicado de forma iterativa: após a identificação e a remoção de um *outlier*, o teste é repetido, até que nenhum outro *outlier* possa ser detectado (TEUNISSEN, 2006). A estatística de teste associada ao DS é dada pelo máximo valor absoluto do vetor de resíduos padronizados (resíduos na mesma escala de valores). Uma vez que o DS é baseado em teste de hipóteses, erros de decisão podem ser cometidos. A descrição das probabilidades de falhas e sucesso associadas com o DS é o ponto essencial da teoria da confiabilidade. O cálculo de tais probabilidades requer informações sobre a função de densidade de probabilidade (f.d.p.) dos resíduos estimados.

Possivelmente devido à falta de recursos computacionais na época, Baarda (1968) se limitou a descrever a teoria da confiabilidade com base nas probabilidades de cometer os erros do tipo I ( $\alpha$ ) e tipo II ( $\beta$ ). O erro tipo I consiste em rejeitar a hipótese nula quando ela é de fato verdadeira e o erro tipo II é a decisão de não rejeitar a hipótese nula quando de fato ela é falsa. Os sistemas de equações sob a hipótese nula são formulados sem a presença de *outlier*, enquanto que a hipótese alternativa descreve um modelo com uma das observações sendo um possível *outlier*. A probabilidade de cometer o erro do tipo I é conhecida como nível de significância do teste ( $\alpha$ ) e seu complemento é o nível de confiança ( $NC = 1 - \alpha$ ), enquanto que o complemento da probabilidade de cometer o erro tipo II ( $\beta$ ) é conhecida como poder do teste ( $\gamma = 1 - \beta$ ). Com base em  $\alpha$  e  $\beta$ , Baarda (1968) introduziu o conceito de menor *outlier* detectável (do inglês, *Minimal Detectable Bias* – MDB). MDB fornece a magnitude do menor *outlier* que pode levar a rejeição da hipótese nula para dados  $\alpha$  e  $\beta$ . O MDB de cada observação descreve a confiabilidade interna do modelo, e sua propagação sobre os parâmetros do modelo, a confiabilidade externa.

Baarda (1968) simplificou o problema da confiabilidade ao desconsiderar a análise de separabilidade. Considerando o fato de que não somente os erros tipo I e II estão presentes nas decisões do DS, Förstner (1981) estendeu a teoria da confiabilidade ao descrever a probabilidade de cometer o erro tipo III ( $\kappa$ ). O erro tipo III, que acontece somente se existir a presença de pelo menos um *outlier* nas medidas, consiste em rejeitar corretamente a hipótese nula, porém identificar (erroneamente) uma observação como contendo *outlier*, enquanto o verdadeiro *outlier* permanece nos dados observados – isto é, aceitar uma hipótese alternativa falsa, enquanto a hipótese alternativa verdadeira não foi aceita (HAWKINS, 1980). Além disso, Förstner (1983) introduziu na análise original de Baarda (1968) o coeficiente de correlação entre as estatísticas de teste de cada par de observações. Entretanto, Förstner (1983) também restringiu a teoria da separabilidade (e, portanto, da confiabilidade) ao considerar apenas duas hipóteses alternativas. A razão disso também se deve possivelmente à limitação dos recursos computacionais disponíveis naquela época.

Yang et al. (2013), utilizando os estudos de Förstner (1983), estende a teoria da confiabilidade de Baarda (1968), mas nesta abordagem, considerando todas as observações envolvidas (múltiplas hipóteses alternativas), ao invés de cada par de observações separadamente. Os autores apresentam uma teoria que estabelece os limites dos erros de decisão do DS sob múltiplas hipóteses alternativas. Entretanto, os mesmos autores apontam que o cálculo de tais limites somente é possível por meio de técnicas de simulações computacionais. Eles estendem a teoria convencional de Baarda (1968), porém com algumas limitações.

Primeiro, fixaram a correlação entre as estatísticas de teste no estudo da relação entre o número de observações e as probabilidades de cometer os tipos de erro (I, II e III). Ao final, Yang et al. (2013) descreve o limiar inferior do poder do teste DS. Além disso, esse limiar inferior estimado se refere somente a uma rodada de aplicação do DS, não descrevendo a real confiabilidade do DS, pois, o mesmo é aplicado na forma iterativa.

Prószyński (2015) estende o MDB por um índice de identificabilidade de *outlier* atribuído a cada observação individual e um índice de falsa identificabilidade como sendo a probabilidade máxima de identificar uma observação como sendo um *outlier*, quando essa de fato não é.

Embora grandes avanços tenham ocorrido nos últimos anos, somente recentemente a teoria da confiabilidade para o DS foi completamente descrita. Rofatto, Matsuoka e Klein (2017) deram o primeiro passo usando a computação inteligente, mais especificamente o método de Monte Carlo (MC), mas não estimaram a confiabilidade interna por observação e, sim, um indicador para toda a rede geodésica. Posteriormente, Rofatto et al. (2020b) estenderam de forma completa a teoria da confiabilidade para o DS, inclusive considerando o procedimento iterativo do mesmo. Primeiro, mostraram que para controlar efetivamente o erro tipo I do DS, as dependências entre os resíduos estimados devem ser consideradas. De forma semelhante a Lehmann (2012), forneceram uma rotina baseada em MC para capturar essas dependências e obter o valor crítico correspondente à probabilidade do erro Tipo I fornecida pelo usuário. Em seguida, na presença de um *outlier*, recorreram ao MC para obter os níveis de probabilidade do DS. Com base nos níveis de probabilidade, eles extraíram dois índices de confiabilidade do DS: *Minimal Identifiable Bias* (MIB) e *Minimal Detectable Bias* (MDB). O MIB expressa a capacidade do DS em identificar e remover corretamente um *outlier*, enquanto o MDB está associado apenas com a correta detecção. Aqui, detecção somente nos informa se a hipótese nula deve ser aceita ou não. A identificação nos diz qual hipótese alternativa (ou seja, qual medida) teria levado à rejeição da hipótese nula. Portanto, ao aplicar o DS podemos ter detecção e não identificação (sabemos que temos *outlier*, mas não sabemos onde ele está), mas sempre que temos identificação obviamente também temos detecção.

Destacamos que até o momento, grande parte dos trabalhos têm sido conduzidos para redes simuladas de nivelamento. Aqui, por outro lado, aplicamos a recente teoria de confiabilidade estendida no contexto de redes baseadas no sistema de posicionamento por satélites GNSS (*Global Navigation Satellite System*), a partir de dados reais coletados em campo. Investigamos as consequências de conhecer as covariâncias entre as componentes das linhas base, uma vez que na fase de planejamento de redes GNSS as covariâncias entre as componentes são normalmente adotadas como nulas e apenas as variâncias das componentes são consideradas. Ressalta-se que o caso de dados com covariâncias não foi contemplado em nenhum estudo prévio até então.

O trabalho está apresentado da seguinte forma: no tópico 2, explicamos os fundamentos envolvido com o DS e seus níveis de probabilidade; no tópico 3, descrevemos o procedimento para obtenção dos níveis de probabilidade do DS e seus correspondentes indicadores de sensibilidade para fins de detecção (MDB) e identificação (MIB) de *outlier*; no tópico 4, fornecemos uma descrição dos dados disponíveis para a realização dos experimentos; em seguida, os resultados são apresentados; e, por fim, destacamos as conclusões do trabalho.

## 2 PROCEDIMENTO PARA IDENTIFICAÇÃO E REMOÇÃO DE *OUTLIER*: DATA SNOOPING

O critério mais adotado nas ciências geodésicas para a solução única de sistemas de equações redundantes e inconsistentes é o método dos mínimos quadrados (MMQ). Entretanto, o MMQ não é um estimador robusto, ou seja, o MMQ é uma solução sensível a *outliers* (ROFATTO et al., 2020b). Portanto, após o ajustamento pelo MMQ ter sido realizado, o procedimento estatístico *Data-Snooping* (DS) é aplicado para a identificação e remoção de possíveis *outliers*.

As etapas envolvidas com o DS são descritas a seguir. Primeiro, os resíduos são determinados no processo de estimação por MMQ, conforme Eq. (1) e Eq. (2):

$$\hat{e} = A\hat{x} - y, \text{ com} \quad (1)$$

$$\hat{x} = (A^T W A)^{-1} (A^T W y) \quad (2)$$

em que  $\hat{e} \in \mathbb{R}^{n \times 1}$  é o vetor dos resíduos estimados,  $A \in \mathbb{R}^{n \times u}$  é a matriz das derivadas parciais em relação aos parâmetros incógnitos do modelo (também conhecida como matriz *design* ou Jacobiana),  $W \in \mathbb{R}^{n \times n}$  é a matriz peso das observações,  $\hat{x} \in \mathbb{R}^{u \times 1}$  é o vetor dos parâmetros estimados e  $y \in \mathbb{R}^{n \times 1}$  é o vetor dos dados observados.

Com base nos resíduos estimados, as estatísticas de teste ( $w_i$ ) para cada observação são calculadas por Eq. (3):

$$w_i = \frac{c_i^T W \hat{e}}{\sqrt{c_i^T W Q_{\hat{e}} W c_i}} \quad \forall i = 1, \dots, n \quad (3)$$

onde  $Q_{\hat{e}} \in \mathbb{R}^{n \times n}$  é matriz covariância dos resíduos e  $c_i$  é um vetor unitário relacionado à  $i$ -ésima observação a ser testada.

Uma vez que os erros observacionais se espalham entre todos os resíduos, o DS tem efeito quando retorna o maior valor absoluto entre os  $w_i$ , isto é Eq. (4):

$$\max-w = \max_{i \in \{1, \dots, n\}} |w_i| \quad (4)$$

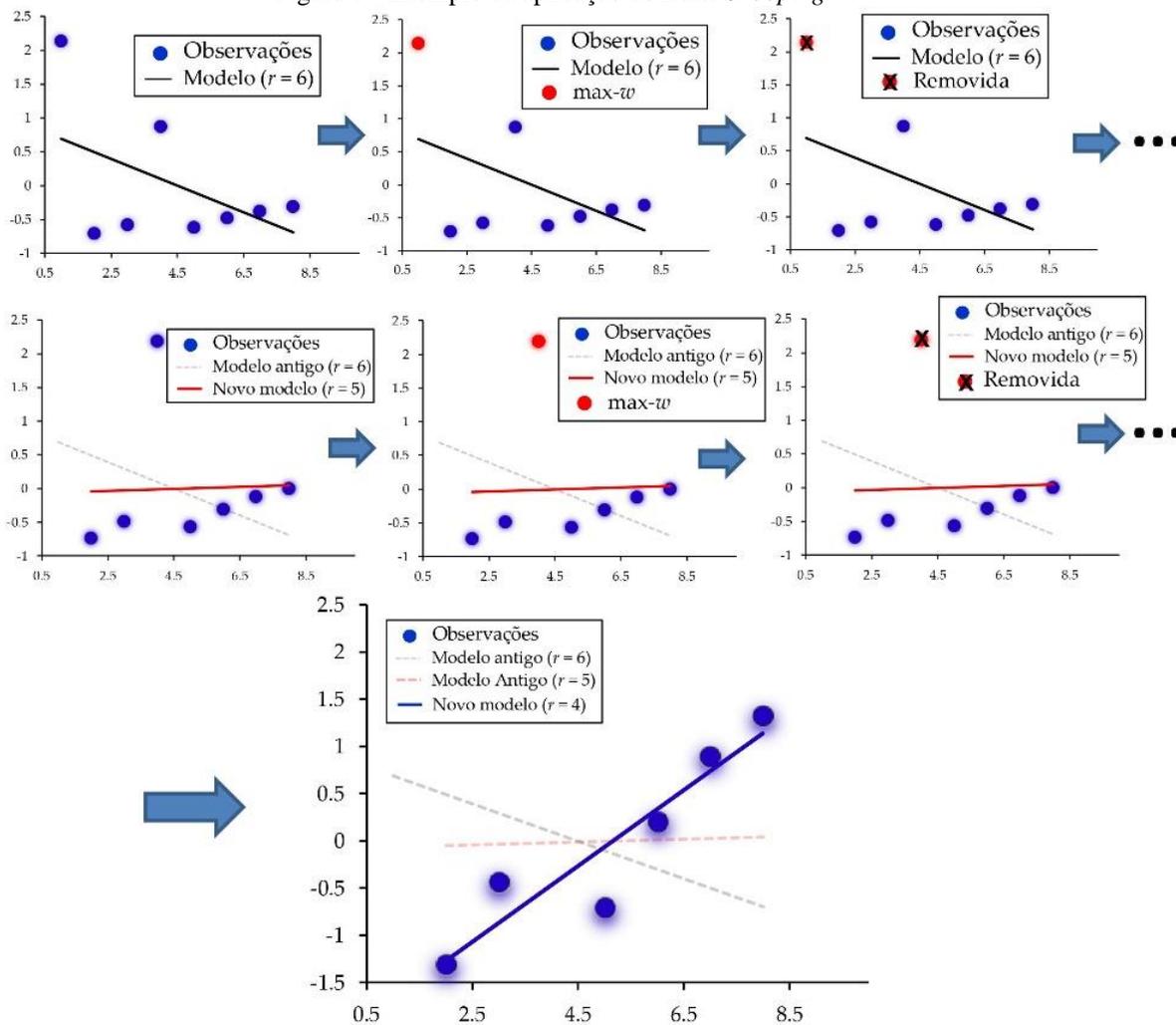
A observação identificada como sendo *outlier* é aquela cuja estatística de teste é máxima e maior que o valor crítico pré-estipulado ( $k$ ), ou seja:

$$\max-w > k \quad (5)$$

Geralmente,  $k$  é obtido com base na distribuição normal bilateral. A escolha do nível de significância e, conseqüentemente, do valor crítico, desempenha um papel fundamental na performance do procedimento DS (NOWEL, 2016; ROFATTO et al., 2020b).

A observação identificada é eliminada do modelo, e se a redundância for suficiente, o procedimento é repetido até que não haja *outliers* nas observações. Nota-se que o MMQ é realizado com  $n - 1$  observações restantes em cada iteração.

Figura 1 - Exemplo da aplicação do *Data-Snooping*.



Fonte: Rofatto (2020).

Um exemplo de aplicação do DS é mostrado na Figura 1, como segue:

- a) primeiro, os parâmetros do modelo são obtidos via MMQ. O modelo inicial é representado pela linha preta. Com base no vetor dos resíduos dos mínimos quadrados, as estatísticas de teste de cada observação, mostradas em azul, são calculadas (denotadas por  $w_i$ );
- b) em seguida, a máxima estatística de teste é encontrada ( $max-w$ ), representado em vermelho;
- c) testa-se sua significância de acordo com o valor crítico. Se a máxima estatística exceder o valor crítico, retira-se o *outlier* identificado do conjunto das observações (marcado em x);
- d) o procedimento é repetido até que todos os *outliers* suspeitos sejam identificados e removidos, em um processo iterativo de estimação, detecção, identificação e remoção do *outlier*.

Observa-se que a redundância global do modelo ( $r = n - u$ ) reduz a  $n - 1$  à medida que um *outlier* é removido. Como consequência, em cada iteração um novo modelo sob a hipótese nula é selecionado e testado contra  $n - 1$  hipóteses alternativas. Portanto, o DS é um processo iterativo de eliminação de *outlier*. Entretanto, por ser uma estratégia que se utiliza de teste estatístico, pode-se cometer erros na tomada de decisão, como segue (ROFATTO et al. 2020b):

- a)  $P_{MD}$  (falha de detecção): probabilidade de não detectar *outliers* (similar a probabilidade de cometer o erro tipo II);
- b)  $P_{WE}$  (exclusão errada): probabilidade de identificar e remover erroneamente uma observação, enquanto o *outlier* permanece no conjunto de dados (ou seja, probabilidade de cometer o erro tipo III);

- c)  $P_{over+}$  (superidentificação positiva): probabilidade de identificar e remover corretamente o *outlier* e outras observações;
- d)  $P_{over-}$  (superidentificação negativa): probabilidade de identificar e remover mais de uma observação como sendo *outlier*, enquanto o *outlier* permanece nos dados observados;
- e)  $P_{ol}$  (sobreposição estatística): probabilidade de detectar simultaneamente duas (ou mais) observações como *outliers*. Isso ocorre quando duas hipóteses alternativas têm a mesma distribuição e não podem ser diferenciadas, ou seja, são inseparáveis e não é possível identificar um *outlier*.

A probabilidade da correta identificação ( $P_{CI}$ ) é o complemento da soma de todos os níveis de falha associados ao DS, isto é:

$$P_{CI} = 1 - (P_{MD} + P_{WE} + P_{over+} + P_{over-} + P_{ol}) \quad (6)$$

### 3 TEORIA ESTENDIDA DA CONFIABILIDADE PARA REDES GEODÉSICAS

Os níveis de probabilidade associados com o DS são analiticamente intratáveis e não podem ser descritos a partir de métodos determinísticos. Por essa razão, recorreremos ao Método de Monte Carlo (ROFATTO et al., 2017)

Entende-se por Monte Carlo qualquer método estatístico que se baseia em amostragens aleatórias massivas para obter resultados numéricos, isto é, repetindo sucessivas simulações um elevado número de vezes, para calcular probabilidades heurísticamente, tal como se, de fato, se registrassem os eventos reais (DALLE, 2005).

Na Geodésia, para fins de controle de qualidade, nós utilizamos o método Monte Carlo para realizar os cálculos dos níveis de probabilidades associados ao teste *Data-Snooping*. Porém, a estatística de teste *Data-Snooping* possui sua própria distribuição de probabilidade e os valores críticos para a estatística de teste não estão disponíveis. Diante dessa limitação, o DS é usualmente aplicado usando valores críticos baseados em um nível de significância para uma única hipótese alternativa, e usa esse mesmo valor para o DS que contém múltiplas hipóteses alternativas. Desta forma, não há um controle real do nível de significância para o DS ao aplica-lo desta maneira.

Neste trabalho, antes de apresentar o método da teoria de confiabilidade estendida, também mostramos uma forma de se aplicar o DS com controle do nível de significância. Para tanto, utilizamos o método Monte Carlo para construir a distribuição de probabilidade do teste DS e obter os valores críticos para a estatística de teste em função do nível de significância desejado (ROFATTO et al., 2020b).

O procedimento para obtenção desse valor crítico para a estatística de teste está descrito na seção 3.1 e o procedimento para obtenção dos níveis de probabilidade associados (teoria da confiabilidade estendida) ao procedimento *Data-Snooping* Iterativo está apresentado na seção 3.2.

#### 3.1 Procedimento para obtenção do valor crítico da estatística *max-w*

Deve-se realizar o cálculo do valor crítico  $\hat{k}$  antes de realizar o procedimento descrito na seção 3.2. O procedimento para estimar o valor crítico de *max-w* para aplicação do DS com controle real do nível de significância é descrito a seguir (ROFATTO et al., 2020b):

- a) deve-se especificar a distribuição de probabilidade da estatística *w-test*. A função densidade de probabilidade atribuída para as estatísticas  $w_i$  sob a hipótese nula  $H_0$  é descrita conforme Eq. (7):

$$(w_1, w_2, w_3, \dots, w_n)^T \sim N(0, R_w) \quad (7)$$

com  $R_w$  sendo a matriz dos coeficientes de correlação entre as estatísticas de teste  $w_i$ . Os elementos da diagonal principal são compostos por 1, e os elementos fora da diagonal as correlações. O

coeficiente de correlação pode assumir valores no intervalo entre -1 e 1;

b) emprega-se o algoritmo *Mersenne Twister* (MATSUMOTO; NISHIMURA, 1998) para gerar sinteticamente as sequências de números aleatórios uniformemente distribuídos no intervalo de 0 a 1, e em seguida é utilizado o método de Transformação *Box-Muller* (BOX; MULLER, 1958) para transformar os números em uma distribuição normal. Assim, a sequência de  $m$  vetores aleatórios de  $w_i$  são gerados de acordo com a Eq.(8);

$$\left[ (w_1, w_2, w_3, \dots, w_n)^{T^{(1)}}, (w_1, w_2, w_3, \dots, w_n)^{T^{(2)}}, \dots, (w_1, w_2, w_3, \dots, w_n)^{T^{(n)}} \right] \quad (8)$$

c) para cada sequência descrita na Eq. 8, deve-se calcular o valor do *max-w*, então teremos:

$$\left( \max_{i \in \{1, \dots, n\}} |w_i|^{(1)}, \max_{i \in \{1, \dots, n\}} |w_i|^{(2)}, \dots, \max_{i \in \{1, \dots, n\}} |w_i|^{(m)} \right) \quad (9)$$

d) os valores da sequência *max-w* dados na etapa c) são classificados em ordem crescente. Essa sequência ordenada, denotada por  $\tilde{w}$ , fornece uma representação discreta da função de densidade cumulativa de *max-w*;

$$\tilde{w}^{(1)} < \tilde{w}^{(2)}, \tilde{w}^{(3)}, \dots, < \tilde{w}^{(m)} \quad (10)$$

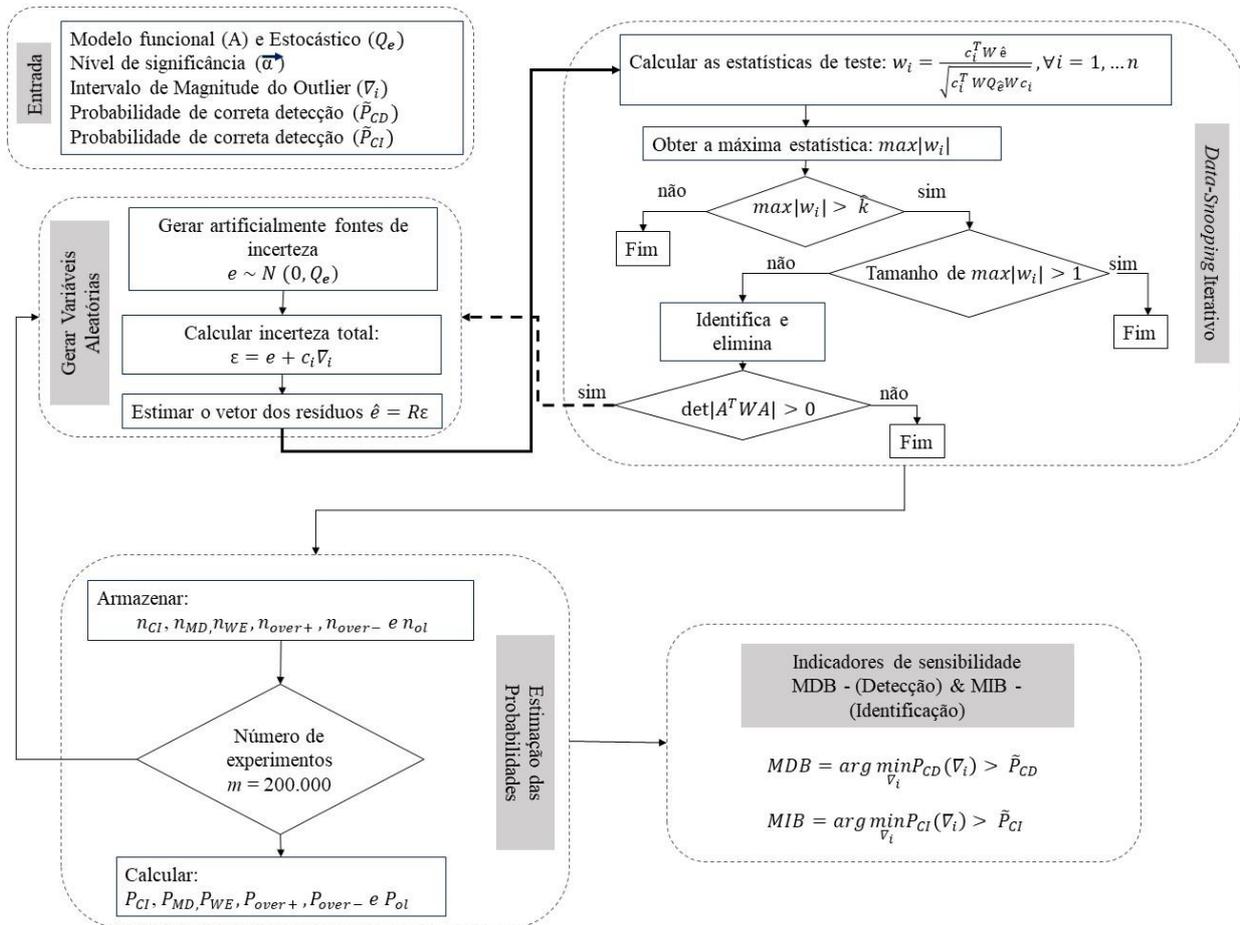
e) por fim, determina-se o valor crítico  $\hat{k}$  em função do nível de significância  $\alpha$  definido pelo usuário, ou seja:

$$\hat{k} = \tilde{w}_{[(1-\alpha) \times m]} \quad (11)$$

### 3.2 Procedimento para estimação dos níveis de probabilidades do procedimento *Data Snooping*

Rofatto et al. (2020b) propõem um método baseado em Monte Carlo (Figura 2) para obter as medidas de confiabilidade de uma rede geodésica.

Figura 2 - Fluxograma do método baseado em Monte Carlo.



Fonte: Rofatto (2020).

Esse método não faz uso de medidas reais coletadas em campo, sendo necessário a matriz design (A), dada pela geometria da rede, e a matriz variância-covariância das observações (Q<sub>e</sub>) como parâmetros de entrada para simulação Monte Carlo.

Considera-se que as medidas coletadas em campo estão contaminadas pelos inevitáveis erros de natureza aleatória tornando imperfeita a relação entre as observações e os parâmetros do modelo. Os erros aleatórios são produzidos de forma artificial por meio do gerador de números aleatórios *Mersenne Twister*, gerando os números uniformemente distribuídos no intervalo [0,1].

Por meio da transformação *Box-Müller* os números aleatórios gerados via *Mersenne Twister* são transformados em uma distribuição normal, pois assumimos que os erros aleatórios seguem distribuição normal. Os erros aleatórios são obtidos sinteticamente conforme Eq. (12).

$$e \sim N(0, Q_e). \tag{12}$$

Porém, as medidas também estão contaminadas por erros de natureza não aleatória, *outliers*. Esses *outliers* são gerados sinteticamente com base no desvio padrão das observações, cada simulação realizada é composta por um vetor de erros aleatórios, dentre os quais uma observação é contaminada por um *outlier* de magnitude  $|3\sigma|$  a  $|9\sigma|$ , com intervalo de  $0,1\sigma$ , como mostra a Eq. (13).

$$\varepsilon = e + c_i V_i \tag{13}$$

Em seguida para que o procedimento *Data-Snooping* seja realizado deve-se definir o vetor dos resíduos ajustados, conforme Eq. (14), contendo os inevitáveis erros aleatórios e os *outliers*.

$$\hat{\epsilon} = R\epsilon \tag{14}$$

em que  $R$  é a matriz redundância obtida conforme a Eq. (15).

$$R = I - A(A^TWA)^{-1}A^TW \tag{15}$$

sendo  $I$  a matriz identidade.

Com base no valor dos resíduos ajustados, dado pela Eq. (14), o procedimento *Data-Snooping* é aplicado. O procedimento é realizado para  $m$  amostras de vetores aleatórios, sendo cada amostra contaminada por um *outlier*.

Com base nos  $m$  experimentos de Monte Carlo, as probabilidades  $P_{CI}$ ,  $P_{MD}$ ,  $P_{WE}$ ,  $P_{over+}$ ,  $P_{over-}$  e  $P_{ol}$  podem ser estimadas, respectivamente por (ROFATTO et al., 2020b):

$$P_{CI} \approx \frac{n_{CI}}{m} \tag{16}$$

$$P_{MD} \approx \frac{n_{MD}}{m} \tag{17}$$

$$P_{WE} \approx \frac{n_{WE}}{m} \tag{18}$$

$$P_{over+} \approx \frac{n_{over+}}{m} \tag{19}$$

$$P_{over-} \approx \frac{n_{over-}}{m} \tag{20}$$

$$P_{ol} \approx \frac{n_{ol}}{m} \tag{21}$$

em que  $n_{CI}$  o número de vezes que um *outlier* é identificado e removido corretamente;  $n_{MD}$  o número de vezes que um *outlier* não é detectado;  $n_{WE}$  o número de vezes que uma observação é erroneamente identificada e removida enquanto o *outlier* permanece no conjunto de dados;  $n_{over+}$  número de vezes que um *outlier* é identificado e removido corretamente, e outras observações são identificadas e removidas de forma errada;  $n_{over-}$  o número de vezes que duas ou mais observações são identificadas e removidas do conjunto de dados observados, enquanto o *outlier* permanece nos dados observados;  $n_{ol}$  número de ocorrências da sobreposição estatística.

A taxa de correta detecção ( $P_{CD}$ ) do *Data-Snooping* depende da soma das probabilidades de correta identificação  $P_{CI}$ , falsa exclusão  $P_{WE}$ , superidentificação com a correta identificação  $P_{over+}$ , superidentificação sem a correta identificação  $P_{over-}$  e sobreposição estatística  $P_{ol}$ , ou seja, a correta detecção é dada conforme Eq. (22):

$$P_{CD} = 1 - P_{MD} = P_{CI} + P_{WE} + P_{over+} + P_{over-} + P_{ol} \tag{22}$$

A partir dos níveis de probabilidade de correta detecção ( $P_{CD}$ ) e correta identificação ( $P_{CI}$ ), pode-se extrair dois índices de sensibilidade do DS (ROFATTO et al., 2020b), conforme Eq. (23) e Eq. (24):

$$MDB = \arg \min_{\forall i} P_{CD}(\forall i) > \tilde{P}_{CD}, \forall i = 1, \dots, n \tag{23}$$

$$MIB = arg \min_{\forall i} P_{CI}(\forall i) > \tilde{P}_{CI}, \forall i = 1, \dots, n \tag{24}$$

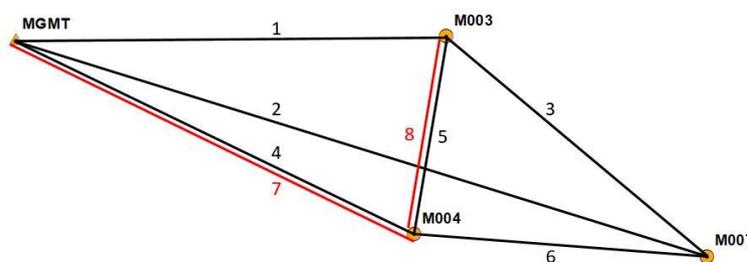
O MDB (*Minimal Detectable Bias*) é definido como o menor valor de um *outlier* que pode ser detectado para uma taxa de correta detecção  $\tilde{P}_{CD}$  definida pelo usuário, Eq. (23). O MDB é um indicador de sensibilidade do DS em termos de detecção. Por outro lado, o MIB (*Minimal Identifiable Bias*) é definido como o menor valor de um *outlier* que pode ser identificado e removido a uma certa taxa de correta identificação  $\tilde{P}_{CI}$ , Eq. (24) – fornecida pelo usuário –, isto é, o MIB é um indicador de sensibilidade do DS para fins de identificação. É importante destacar que a detecção apenas nos informa a existência de um possível *outlier*, enquanto que a identificação nos diz qual observação é um *outlier*.

## 4 MATERIAL E MÉTODOS

### 4.1 Dados Disponíveis

A rede GNSS (Figura 3) utilizada para realizar o experimento é composta por um ponto de coordenadas conhecidas (MGMT) e três pontos desconhecidos (M003, M004 e M007). As componentes da linha base  $\Delta X$ ,  $\Delta Y$  e  $\Delta Z$ , foram estimadas a partir das soluções individuais de cada vetor linha base independente, proporcionada pelo processamento das duplas diferenças da fase da onda portadora L1 no modo estático.

Figura 3 - Rede GNSS.



Fonte: Os autores (2021).

As linhas base (Tabela 1) que compõem a rede foram planejadas de modo a manter a redundância da rede, mantendo todos os pontos com no mínimo três conexões, para ser possível a identificação de pelo menos um *outlier*, conforme recomendado em Klein et al. (2017). Ressalta-se que linhas base repetidas, como as apresentadas em vermelho na Figura 1, melhoram a confiabilidade da rede GNSS (COLLISCHONN et al., 2015). A rede possui uma correlação média entre os resíduos de 62,98%, com mínimo de 61,37% e máximo de 65,59%. O número de redundância local se situa entre 0,4835 e 0,7901.

Tabela 1 - Linhas de base da rede GNSS.

| Vetor | Linha base  | Distância (km) | Data de Rastreamento | Intervalo de Rastreamento (horas) |
|-------|-------------|----------------|----------------------|-----------------------------------|
| 1     | MGMT – M003 | 2,70           | 16/10/2019           | 5                                 |
| 2     | MGMT – M007 | 4,54           | 16/10/2019           | 5                                 |
| 3     | M003 – M007 | 2,14           | 05/09/2016           | 7                                 |
| 4     | M004 – MGMT | 2,78           | 16/10/2019           | 5                                 |
| 5     | M004 – M003 | 1,26           | 06/09/2019           | 5                                 |
| 6     | M004 – M007 | 1,84           | 12/09/2016           | 7                                 |
| 7     | M004 – MGMT | 2,78           | 06/09/2019           | 5                                 |
| 8     | M004 – M003 | 1,26           | 28/08/2016           | 7                                 |

Fonte: Os autores (2021).

A matriz variância-covariância (MVC) das componentes da linha base foi obtida por meio do processamento realizado no software *TopConTools v.8.2.3* (Tabela 2). A MVC é um dos elementos necessários para aplicação do método aqui proposto – como visto no tópico 3.2.

Tabela 2 - Variância e Covariâncias de  $\Delta X$ ,  $\Delta Y$  e  $\Delta Z$  para cada vetor linha base.

| Vetor | $\sigma_{\Delta X}$ (mm <sup>2</sup> ) | $\sigma_{\Delta Y}$ (mm <sup>2</sup> ) | $\sigma_{\Delta Z}$ (mm <sup>2</sup> ) | $\sigma_{\Delta X \Delta Y}$ (mm <sup>2</sup> ) | $\sigma_{\Delta X \Delta Z}$ (mm <sup>2</sup> ) | $\sigma_{\Delta Y \Delta Z}$ (mm <sup>2</sup> ) |
|-------|--|--|--|---|---|---|
| 1     | 256,0                                  | 265,6                                  | 94,0                                   | -180,0  | -79,9   | 70,9  |
| 2     | 552,2                                  | 795,2                                  | 225,0                                  | -480,0  | -160,0  | 157,0   |
| 3     | 306,2                                  | 316,8                                  | 102,0                                  | -230,0  | -87,5   | 83,7  |
| 4     | 278,8                                  | 316,8                                  | 118,8                                  | -210,0  | -91,5   | 95,4  |
| 5     | 176,8                                  | 219,04                                 | 68,8                                   | -140,0  | -54,2   | 51,2  |
| 6     | 262,4                                  | 289,0                                  | 96,0                                   | -200,0  | -83,4   | 80,2  |
| 7     | 484,0                                  | 538,2                                  | 169,0                                  | -370,0  | -140,0  | 142,0   |
| 8     | 228,0                                  | 252,8                                  | 82,8                                   | -170,0  | -67,3   | 65,3  |

Fonte: Os autores (2021).

## 4.2 Delineamento Experimental e Procedimentos Específicos

As taxas de correta identificação e detecção foram obtidas para cada cenário apresentado no Quadro 1. No primeiro cenário (A), consideramos as covariâncias não-nulas entre os componentes das linhas base para  $\alpha = 0,001$  (0,1%). No segundo cenário (B), atribuímos covariâncias nulas entre as componentes para  $\alpha = 0,001$  (0,1%). O terceiro (C) e o quarto (D) cenário são os mesmos modelos estocásticos definidos em (A) e (B), respectivamente, entretanto ambos com  $\alpha = 0,1$  (10%).

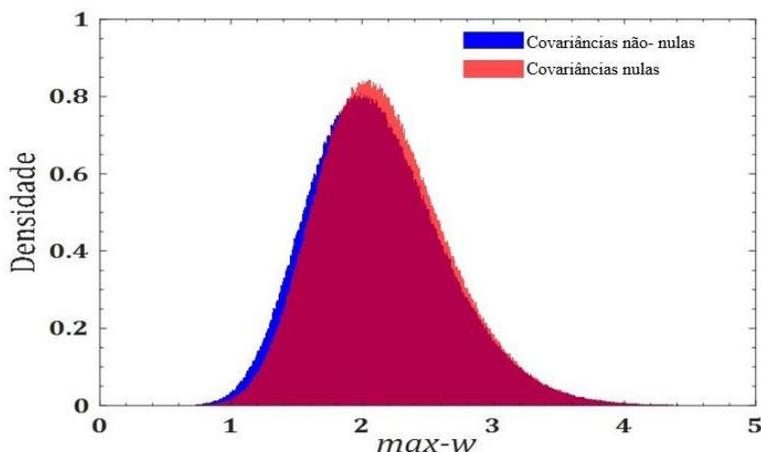
Quadro 1 - Delineamento Experimental.

| Cenários | Modelo Estocástico     | Nível de significância  |
|----------|------------------------|-------------------------|
| A        | Covariâncias não-nulas | $\alpha = 0,001$ (0,1%) |
| B        | Covariâncias nulas     | $\alpha = 0,001$ (0,1%) |
| C        | Covariâncias não-nulas | $\alpha = 0,1$ (10%)    |
| D        | Covariâncias nulas     | $\alpha = 0,1$ (10%)    |

Fonte: Os autores (2021).

Os valores críticos ( $\hat{k}$ ) encontrados foram obtidos conforme o procedimento descrito na seção 3.1, como segue: 4,08 ( $\alpha = 0,1\%$ ) e 2,80 ( $\alpha = 10\%$ ) para o caso das covariâncias não-nulas; e 4,11 ( $\alpha = 0,1\%$ ) e 2,82 ( $\alpha = 10\%$ ) para o caso das covariâncias nulas. De acordo com o procedimento descrito na seção 3.1, obtivemos uma representação discreta da função de densidade cumulativa de  $max-w$  (Figura 4) para ambos os cenários (covariâncias não-nulas e nulas). Nota-se pela Figura 4 que a distribuição de  $max-w$  para o caso de covariâncias nulas é ligeiramente diferente em relação ao caso das covariâncias não-nulas. Essa é a razão dos valores críticos não serem significativamente diferentes em A e B, como em C e D.

Figura 4 - Histograma de densidade do  $max-w$  para os cenários com covariâncias não-nulas (azul) e nulas (vermelho).



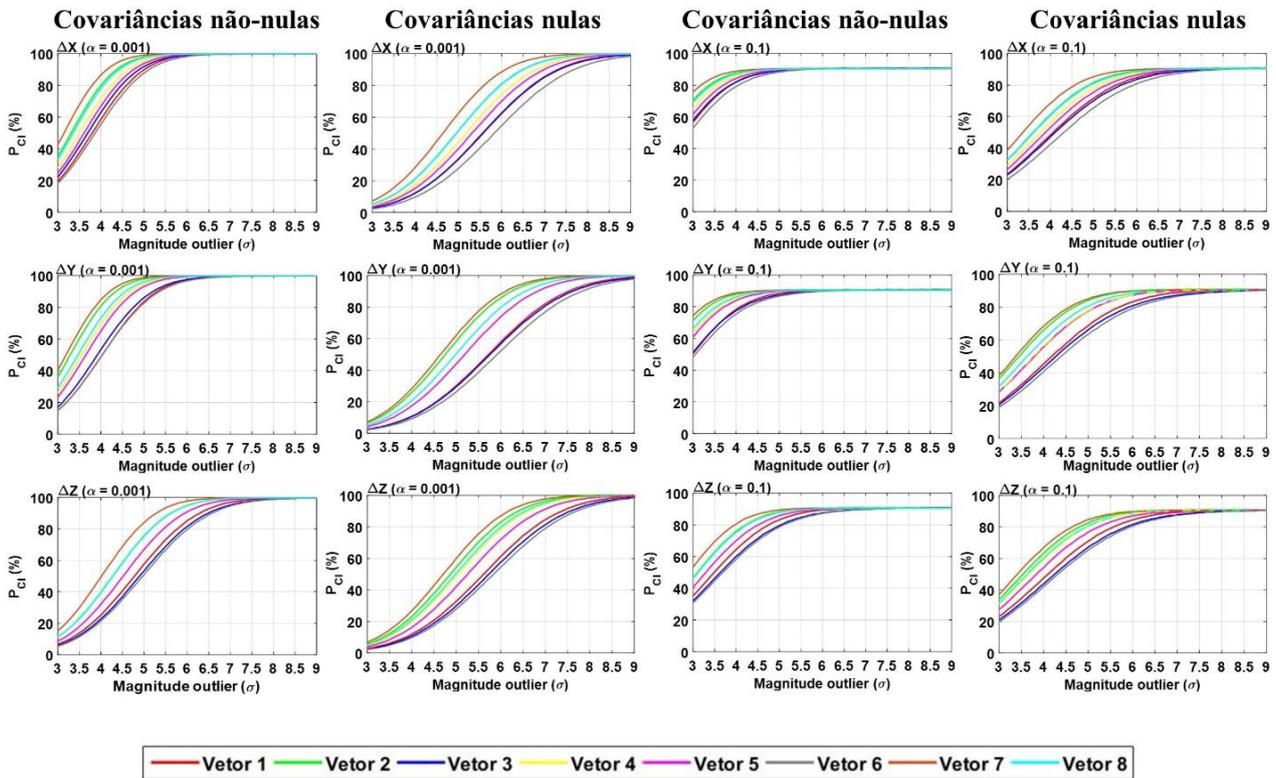
Fonte: Os autores (2021).

## 5 DESCRIÇÃO DOS NÍVEIS DE PROBABILIDADE DO DS PARA OS CENÁRIOS COM COVARIÂNCIAS NÃO-NULAS E COVARIÂNCIAS NULAS

Os níveis de probabilidade associados ao procedimento DS foram computados de acordo com o

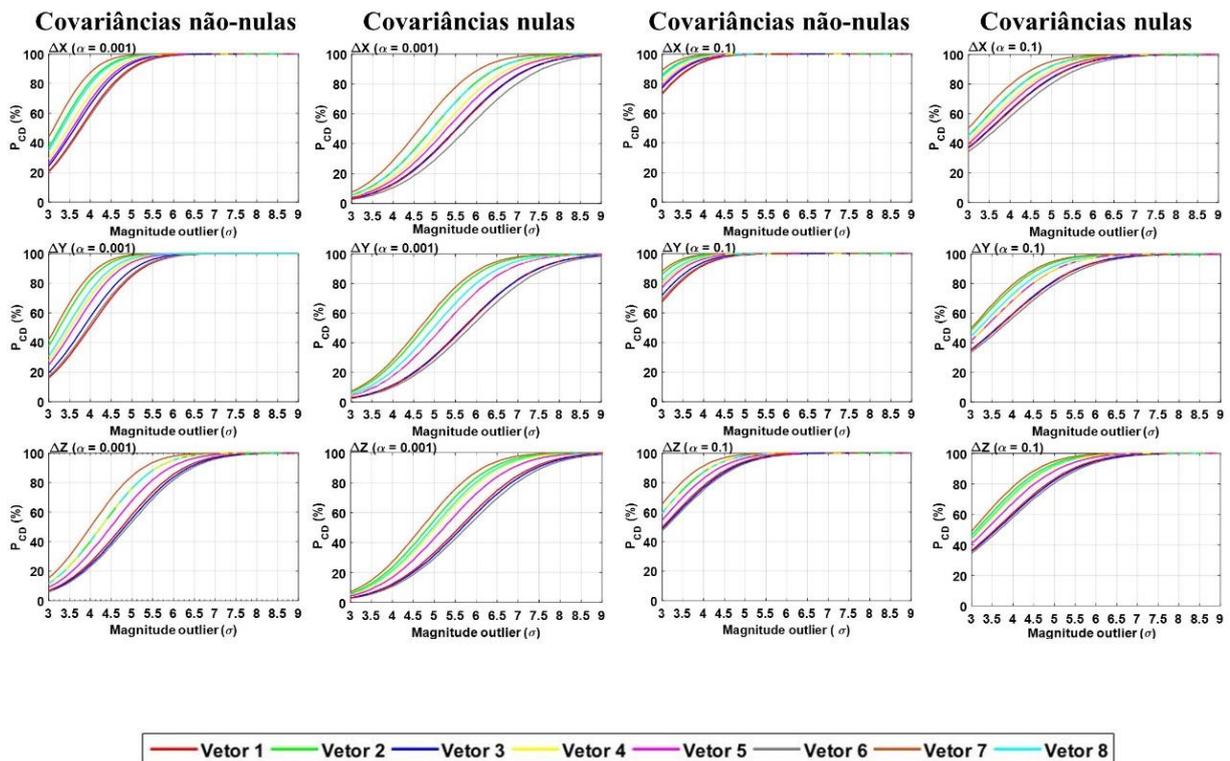
procedimento descrito na seção 3.2 (Figura 2). Embora os valores críticos por MC não apresentaram diferenças, o cenário com covariâncias não-nulas possui taxas de correta identificação e detecção maiores que o cenário com covariâncias nulas (Figuras 5 e 6).

Figura 5 -  $P_{CI}$  para os cenários com covariâncias nulas e não-nulas para  $\alpha = 0,001$  e  $\alpha = 0,1$ .



Fonte: Os autores (2021).

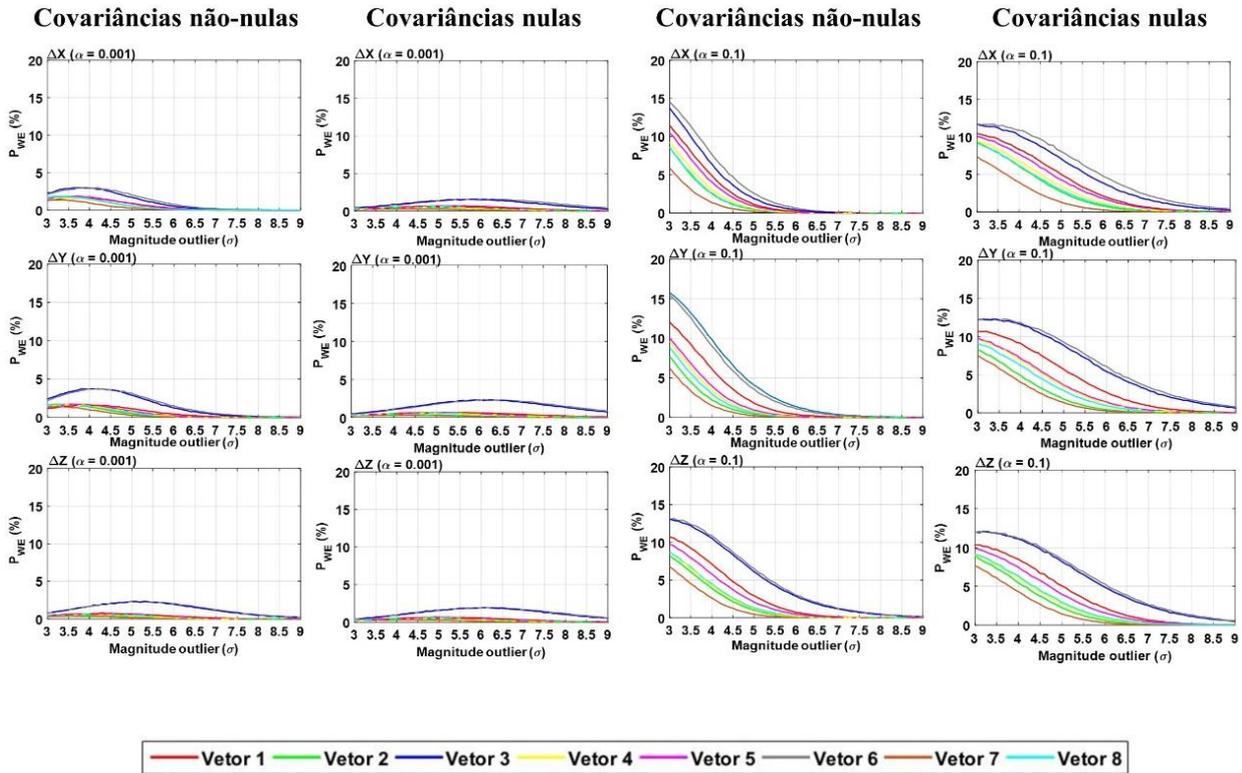
Figura 6 -  $P_{CD}$  para os cenários com covariâncias nulas e não-nulas para  $\alpha = 0,001$  e  $\alpha = 0,1$ .



Fonte: Os autores (2021).

As probabilidades do erro tipo III são apresentados na Figura 7. Observa-se baixa probabilidade do erro tipo III ( $P_{WE} < 5\%$ ) para o  $\alpha = 0,001$ , tanto para o cenário com covariâncias não-nulas quanto o de covariâncias nulas. Entretanto, nota-se uma alta taxa do erro tipo III ( $P_{WE} > 10\%$ ) para o caso em que  $\alpha = 0,1$ .

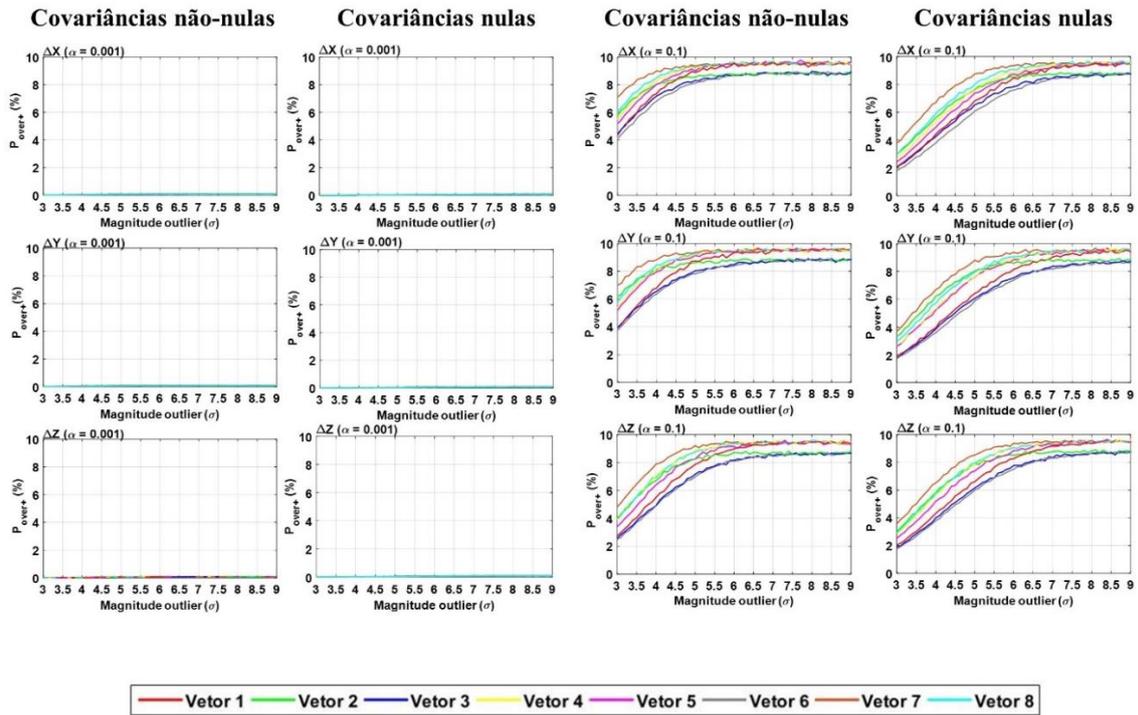
Figura 7 -  $P_{WE}$  para os cenários com covariâncias nulas e não-nulas para  $\alpha = 0,001$  e  $\alpha = 0,1$ .



Fonte: Os autores (2021).

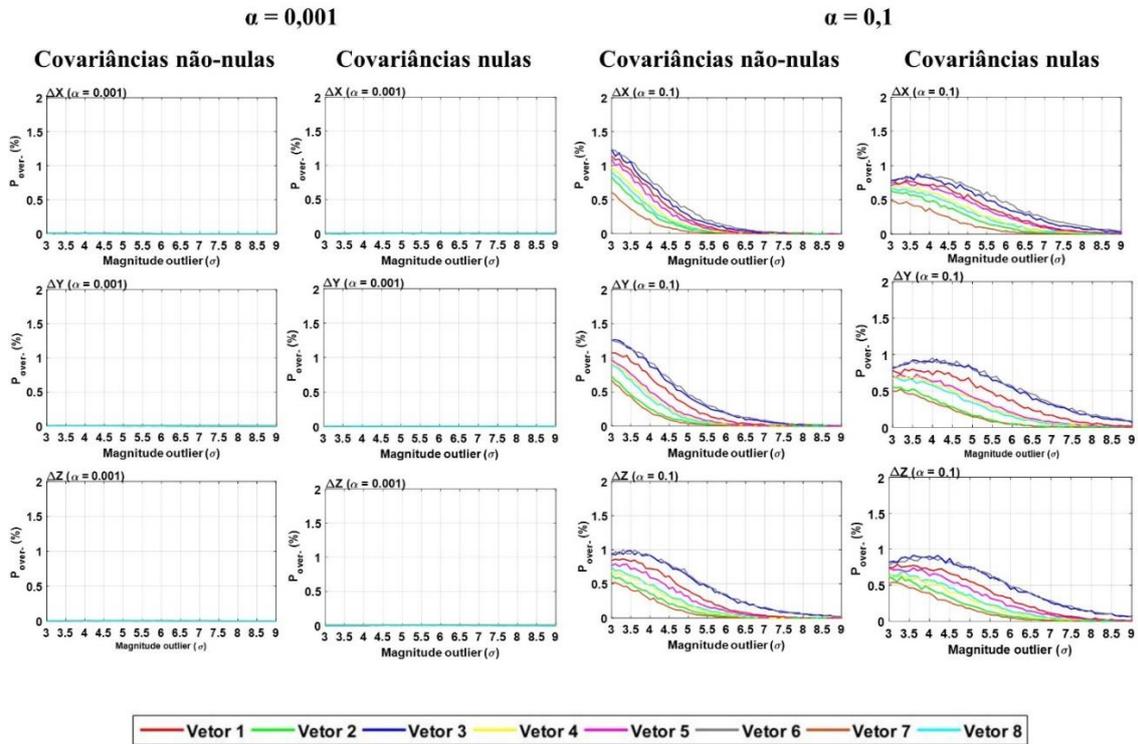
A seguir são apresentadas as classes de superidentificação positiva ( $P_{over+}$ ), Figura 8, e superidentificação negativa ( $P_{over-}$ ), Figura 9. Nota-se que a probabilidade de superidentificação positiva ( $P_{over+}$ ) para  $\alpha = 0,001$  é praticamente nula em ambos os cenários. Entretanto, o aumento do nível de significância ( $\alpha = 0,1$ ), aumentou a taxa de  $P_{over+}$  em ambos os cenários. Além disso, observa-se que quanto maior a magnitude do outlier, maior é a taxa de  $P_{over+}$ . No que diz respeito a  $P_{over-}$ , as taxas foram praticamente nulas em todos os cenários ( $P_{over-} < 1,4\%$ ). Como a classe de sobreposição estatística foi praticamente nula em todos os cenários, ela não foi mostrada aqui.

Figura 8 -  $P_{over+}$  para os cenários com covariâncias nulas e não-nulas para  $\alpha = 0,001$  e  $\alpha = 0,1$



Fonte: Os autores (2021).

Figura 9 -  $P_{over-}$  para os cenários com covariâncias nulas e não-nulas para  $\alpha = 0,001$  e  $\alpha = 0,1$ .



Fonte: Os autores (2021).

As probabilidades de correta identificação ( $P_{CI}$ ) e correta detecção ( $P_{CD}$ ) apresentam o mesmo comportamento em relação ao nível de significância e a magnitude do *outlier*, ou seja, conforme aumenta o nível de significância e a magnitude do *outlier*, maiores são as taxas de correta identificação e correta detecção. A rede GNSS apresenta maiores probabilidades de  $P_{WE}$ ,  $P_{over+}$  e  $P_{over-}$  para o nível de significância  $\alpha = 0,1$  (10%). Porém, o comportamento das classes é o mesmo, ou seja, níveis de significância maiores possuem

maiores taxas de probabilidades das classes:  $P_{WE}$ ,  $P_{over+}$  e  $P_{over-}$ .

## 6 AVALIAÇÃO DAS TAXAS DE SUCESSO DO DS ENTRE OS CENÁRIOS COM COVARIÂNCIAS NÃO-NULAS E COVARIÂNCIAS NULAS

Os indicadores de sensibilidade para identificação (MIB) e detecção (MDB) foram obtidos para as taxas iguais de correta identificação e detecção, ou seja,  $\tilde{P}_{CI} = \tilde{P}_{CD} = 80\%$  (Tabela 3 e Tabela 4). As diferenças entre MDB e MIB (Tabela 5) tiveram valores distintos.

Tabela 3 - MDB para  $\tilde{P}_{CD} = 80\%$ .

| Vetor | Covariâncias não-nulas ( $\sigma$ )<br>$\alpha = 0,001$ |            |            | Covariâncias nulas ( $\sigma$ )<br>$\alpha = 0,001$ |            |            | Covariâncias não-nulas ( $\sigma$ )<br>$\alpha = 0,1$ |            |            | Covariâncias nulas ( $\sigma$ )<br>$\alpha = 0,1$ |            |            |
|-------|---|------------|------------|---|------------|------------|---|------------|------------|---|------------|------------|
|       | $\Delta X$  | $\Delta Y$ | $\Delta Z$ | $\Delta X$  | $\Delta Y$ | $\Delta Z$ | $\Delta X$  | $\Delta Y$ | $\Delta Z$ | $\Delta X$  | $\Delta Y$ | $\Delta Z$ |
| 1     | 4,6   | 4,9        | 5,8        | 6,7   | 6,9        | 6,7        | 3,3   | 3,5        | 4,1        | 4,8   | 5,0        | 5,0        |
| 2     | 4,0   | 4,0        | 5,2        | 6,0   | 5,8        | 5,9        | 3,0   | 3,0        | 3,7        | 4,3   | 4,2        | 4,2        |
| 3     | 4,4   | 4,7        | 5,9        | 6,7   | 6,9        | 6,8        | 3,1   | 3,3        | 4,2        | 4,8   | 4,9        | 4,9        |
| 4     | 4,2   | 4,3        | 5,2        | 6,2   | 6,2        | 6,1        | 3,0   | 3,1        | 3,7        | 4,5   | 4,5        | 4,4        |
| 5     | 4,4   | 4,4        | 5,5        | 6,4   | 6,2        | 6,3        | 3,1   | 3,1        | 3,9        | 4,6   | 4,5        | 4,6        |
| 6     | 4,6   | 4,8        | 6,0        | 7,0   | 7,0        | 7,0        | 3,3   | 3,4        | 4,3        | 5,0   | 5,0        | 5,0        |
| 7     | 3,8   | 3,9        | 4,9        | 5,6   | 5,6        | 5,7        | 3,0   | 3,0        | 3,5        | 4,1   | 4,0        | 4,1        |
| 8     | 4,1   | 4,2        | 5,2        | 6,0   | 6,0        | 6,0        | 3,0   | 3,0        | 3,7        | 4,3   | 4,3        | 4,3        |

Fonte: Os autores (2021).

Tabela 4 - MIB para  $\tilde{P}_{CI} = 80\%$ .

| Vetor | Covariâncias não-nulas ( $\sigma$ )<br>$\alpha = 0,001$ |            |            | Covariâncias nulas ( $\sigma$ )<br>$\alpha = 0,001$ |            |            | Covariâncias não-nulas ( $\sigma$ )<br>$\alpha = 0,1$ |            |            | Covariâncias nulas ( $\sigma$ )<br>$\alpha = 0,1$ |            |            |
|-------|---|------------|------------|---|------------|------------|---|------------|------------|---|------------|------------|
|       | $\Delta X$  | $\Delta Y$ | $\Delta Z$ | $\Delta X$  | $\Delta Y$ | $\Delta Z$ | $\Delta X$  | $\Delta Y$ | $\Delta Z$ | $\Delta X$  | $\Delta Y$ | $\Delta Z$ |
| 1     | 4,6   | 4,9        | 5,8        | 6,7   | 6,9        | 6,8        | 3,9   | 4,1        | 4,8        | 5,6   | 5,8        | 5,6        |
| 2     | 4,1   | 4,0        | 5,2        | 6,0   | 5,8        | 5,9        | 3,4   | 3,4        | 4,3        | 4,9   | 4,7        | 4,9        |
| 3     | 4,5   | 4,8        | 6,0        | 6,7   | 7,0        | 6,9        | 3,9   | 4,2        | 5,1        | 5,7   | 6,0        | 5,9        |
| 4     | 4,3   | 4,3        | 5,2        | 6,2   | 6,2        | 6,1        | 3,6   | 3,7        | 4,3        | 5,2   | 5,2        | 5,0        |
| 5     | 4,4   | 4,5        | 5,5        | 6,4   | 6,3        | 6,3        | 3,7   | 3,8        | 4,6        | 5,4   | 5,2        | 5,3        |
| 6     | 4,7   | 4,9        | 6,1        | 7,0   | 7,2        | 7,1        | 4,1   | 4,3        | 5,1        | 6,0   | 6,1        | 6,0        |
| 7     | 3,8   | 3,9        | 4,9        | 5,6   | 5,6        | 5,7        | 3,2   | 3,3        | 4,0        | 4,6   | 4,6        | 4,7        |
| 8     | 4,1   | 4,2        | 5,2        | 6,0   | 6,1        | 6,1        | 3,5   | 3,6        | 4,3        | 5,0   | 5,0        | 5,0        |

Fonte: Os autores (2021).

Tabela 5 - Diferenças entre MDB e MIB para  $\tilde{P}_{CI} = \tilde{P}_{CD} = 80\%$ .

| Vetor         | Covariâncias não-nulas ( $\sigma$ )<br>$\alpha = 0,001$ |            |            | Covariâncias nulas ( $\sigma$ )<br>$\alpha = 0,001$ |            |            | Covariâncias não-nulas ( $\sigma$ )<br>$\alpha = 0,1$ |            |            | Covariâncias nulas ( $\sigma$ )<br>$\alpha = 0,1$ |            |            |
|---------------|---|------------|------------|---|------------|------------|---|------------|------------|---|------------|------------|
|               | $\Delta X$  | $\Delta Y$ | $\Delta Z$ | $\Delta X$  | $\Delta Y$ | $\Delta Z$ | $\Delta X$  | $\Delta Y$ | $\Delta Z$ | $\Delta X$  | $\Delta Y$ | $\Delta Z$ |
| 1             | 0,0   | 0,0        | 0,0        | 0,0   | 0,0        | -0,1       | -0,6  | -0,6       | -0,7       | -0,8  | -0,8       | -0,6       |
| 2             | -0,1  | 0,0        | 0,0        | 0,0   | 0,0        | 0,0        | -0,4  | -0,4       | -0,6       | -0,6  | -0,5       | -0,7       |
| 3             | -0,1  | -0,1       | -0,1       | 0,0   | -0,1       | -0,1       | -0,8  | -0,9       | -0,9       | -0,9  | -1,1       | -1,0       |
| 4             | -0,1  | 0,0        | 0,0        | 0,0   | 0,0        | 0,0        | -0,6  | -0,6       | -0,6       | -0,7  | -0,7       | -0,6       |
| 5             | 0,0   | -0,1       | 0,0        | 0,0   | -0,1       | 0,0        | -0,6  | -0,7       | -0,7       | -0,8  | -0,7       | -0,7       |
| 6             | -0,1  | -0,1       | -0,1       | 0,0   | -0,2       | -0,1       | -0,8  | -0,9       | -0,8       | -1,0  | -1,1       | -1,0       |
| 7             | 0,0   | 0,0        | 0,0        | 0,0   | 0,0        | 0,0        | -0,2  | -0,3       | -0,5       | -0,5  | -0,6       | -0,6       |
| 8             | 0,0   | 0,0        | 0,0        | 0,0   | -0,1       | -0,1       | -0,5  | -0,6       | -0,6       | -0,7  | -0,7       | -0,7       |
| Média         | 0,0   | 0,0        | 0,0        | 0,0   | -0,1       | 0,0        | -0,6  | -0,6       | -0,7       | -0,8  | -0,8       | -0,7       |
| Desvio-Padrão | 0,1   | 0,1        | 0,0        | 0,0   | 0,1        | 0,1        | 0,2   | 0,2        | 0,1        | 0,2   | 0,2        | 0,2        |
| Máximo        | 0   | 0          | 0          | 0   | 0          | 0          | -0,2  | -0,3       | -0,5       | -0,5  | -0,5       | -0,6       |
| Mínimo        | -0,1  | -0,1       | -0,1       | 0   | -0,2       | -0,1       | -0,8  | -0,9       | -0,9       | -1  | -1,1       | -1         |

Fonte: Os autores (2021).

Para  $\alpha = 0,001$ , independente das componentes  $\Delta X$ ,  $\Delta Y$  e  $\Delta Z$  e do fato de covariâncias nulas e não-nulas, elas variaram no intervalo de  $0,0\sigma$  a  $0,2\sigma$  – ou seja, são na prática de mesmo valor. Já para  $\alpha = 0,1$  o

intervalo de variação foi de  $0,2\sigma$  a  $1,1\sigma$  – essa variação se deve à taxa de erro tipo III ( $P_{WE}$ ) e superidentificação positiva ( $P_{over+}$ ), conforme mostra as Figuras 6 e 7. Os valores negativos na tabela são explicados pelo fato de o MIB ser sempre maior ou igual ao MDB, pois identificar é mais difícil do que apenas detectar.

As diferenças absolutas dos MIBs entre os cenários de covariâncias nulas e não-nulas para  $\tilde{P}_{CI} = 80\%$  – denotadas por  $\Delta MIB$  – variaram no intervalo de  $0,7\sigma$  a  $2,3\sigma$  para  $\alpha = 0,001$  e  $0,6\sigma$  a  $1,9\sigma$  para  $\alpha = 0,1$  (Tabela 6). No caso do MDB – diferenças denotadas por  $\Delta MDB$  –, as diferenças absolutas variaram no intervalo de  $0,7\sigma$  a  $2,4\sigma$  para  $\alpha = 0,001$ , e  $0,7\sigma$  a  $1,7\sigma$  para  $\alpha = 0,1$ . Portanto, a presença das covariâncias reduziu tanto o MIB quanto o MDB, praticamente na mesma proporção (na média  $\sim 30\%$  para  $\Delta X$  e  $\Delta Y$ , e  $\sim 14\%$  para  $\Delta Z$ ). Essa melhoria foi praticamente igual para ambos os níveis de significância considerados ( $\alpha = 0,001$  e  $\alpha = 0,1$ ).

Por outro lado, o aumento do nível de significância reduziu as diferenças entre os cenários de covariâncias nulas e não-nulas (Tabela 7), com redução entre  $0,1\sigma$  e  $0,5\sigma$  para o MIB e  $0\sigma$  a  $0,7\sigma$  para o MDB. Em termos relativos, o aumento do nível de significância de  $\alpha = 0,001$  para  $\alpha = 0,1$  reduziu o MIB em média de  $\sim 20\%$  para  $\Delta X$  e  $\Delta Y$ , e  $\sim 16\%$  para  $\Delta Z$ , e no que tange ao MDB  $\sim 28\%$  para  $\Delta X$  e  $\Delta Y$ , e  $\sim 23\%$  para  $\Delta Z$ . Em termos de MIB, a redução média corresponde à  $0,4\sigma$  para  $\Delta X$  e  $\Delta Y$ , e  $0,1\sigma$  para  $\Delta Z$ , ou seja,  $6,6\text{mm}$  para  $\Delta X$ ,  $8,0\text{mm}$  para  $\Delta Y$  e  $1,5\text{mm}$  para  $\Delta Z$ , enquanto para o MDB essa redução média corresponde à  $0,6\sigma$  para  $\Delta X$  e  $\Delta Y$ , e  $0,2\sigma$  para  $\Delta Z$ , ou seja,  $10,5\text{mm}$  para  $\Delta X$ ,  $10,3\text{mm}$  para  $\Delta Y$  e  $2,1\text{mm}$  para  $\Delta Z$ . Lembrando que o MIB e MDB representam a menor tendência identificável e detectável, respectivamente, e, portanto, não podem ser confundidos com indicadores de precisão.

Tabela 6 - Diferenças entre os cenários de covariâncias nulas e não-nulas para o MIB e MDB em função do nível de significância ( $\tilde{P}_{CI} = \tilde{P}_{CD} = 80\%$ ).

| Vetor         | $\Delta MIB (\sigma)$ (diferenças absolutas) |            |            |                |            |            | $\Delta MDB (\sigma)$ (diferenças absolutas) |            |            |                |            |            |
|---------------|--|------------|------------|----------------|------------|------------|--|------------|------------|----------------|------------|------------|
|               | $\alpha = 0,001$                             |            |            | $\alpha = 0,1$ |            |            | $\alpha = 0,001$                             |            |            | $\alpha = 0,1$ |            |            |
|               | $\Delta X$                                   | $\Delta Y$ | $\Delta Z$ | $\Delta X$     | $\Delta Y$ | $\Delta Z$ | $\Delta X$                                   | $\Delta Y$ | $\Delta Z$ | $\Delta X$     | $\Delta Y$ | $\Delta Z$ |
| 1             | 2,1  | 2,0        | 1,0        | 1,7            | 1,7        | 0,8        | 2,1  | 2,0        | 0,9        | 1,5            | 1,5        | 0,9        |
| 2             | 1,9  | 1,8        | 0,7        | 1,5            | 1,3        | 0,6        | 2,0  | 1,8        | 0,7        | 1,3            | 1,2        | 0,5        |
| 3             | 2,2  | 2,2        | 0,9        | 1,8            | 1,8        | 0,8        | 2,3  | 2,2        | 0,9        | 1,7            | 1,6        | 0,7        |
| 4             | 1,9  | 1,9        | 0,9        | 1,6            | 1,5        | 0,7        | 2,0  | 1,9        | 0,9        | 1,5            | 1,4        | 0,7        |
| 5             | 2,0  | 1,8        | 0,8        | 1,7            | 1,4        | 0,7        | 2,0  | 1,8        | 0,8        | 1,5            | 1,4        | 0,7        |
| 6             | 2,3  | 2,3        | 1,0        | 1,9            | 1,8        | 0,9        | 2,4  | 2,2        | 1,0        | 1,7            | 1,6        | 0,7        |
| 7             | 1,8  | 1,7        | 0,8        | 1,4            | 1,3        | 0,7        | 1,8  | 1,7        | 0,8        | 1,1            | 1,0        | 0,6        |
| 8             | 1,9  | 1,9        | 0,9        | 1,5            | 1,4        | 0,7        | 1,9  | 1,8        | 0,8        | 1,3            | 1,3        | 0,6        |
| Média         | 2,0  | 2,0        | 0,9        | 1,6            | 1,5        | 0,7        | 2,0  | 1,9        | 0,9        | 1,5            | 1,4        | 0,7        |
| Desvio-Padrão | 0,2  | 0,2        | 0,1        | 0,2            | 0,2        | 0,1        | 0,2  | 0,2        | 0,1        | 0,2            | 0,1        | 0,1        |
| Máximo        | 2,3  | 2,3        | 1          | 1,9            | 1,8        | 0,9        | 2,4  | 2,2        | 1          | 1,7            | 1,6        | 0,9        |
| Mínimo        | 1,8  | 1,7        | 0,7        | 1,4            | 1,3        | 0,6        | 1,8  | 1,7        | 0,7        | 1,1            | 1          | 0,5        |

Fonte: Os autores (2021).

Tabela 7 - Diferenças entre  $\Delta MIB (\alpha = 0,001)$  e  $\Delta MIB (\alpha = 0,1)$  para  $\tilde{P}_{CI} = 80\%$  e diferenças entre  $\Delta MDB (\alpha = 0,001)$  e  $\Delta MDB (\alpha = 0,1)$  para  $\tilde{P}_{CD} = 80\%$

| Vetor           | $\Delta MIB (\alpha = 0,001) - \Delta MIB (\alpha = 0,1) (\sigma)$ |            |            | $\Delta MDB (\alpha = 0,001) - \Delta MDB (\alpha = 0,1) (\sigma)$ |            |            |
|-----------------|--|------------|------------|--|------------|------------|
|                 | $\Delta X$   | $\Delta Y$ | $\Delta Z$ | $\Delta X$   | $\Delta Y$ | $\Delta Z$ |
| 1               | 0,4  | 0,3        | 0,2        | 0,6  | 0,5        | 0,0        |
| 2               | 0,4  | 0,5        | 0,1        | 0,7  | 0,6        | 0,2        |
| 3               | 0,4  | 0,4        | 0,1        | 0,6  | 0,6        | 0,2        |
| 4               | 0,3  | 0,4        | 0,2        | 0,5  | 0,5        | 0,2        |
| 5               | 0,3  | 0,4        | 0,1        | 0,5  | 0,4        | 0,1        |
| 6               | 0,4  | 0,5        | 0,1        | 0,7  | 0,6        | 0,3        |
| 7               | 0,4  | 0,4        | 0,1        | 0,7  | 0,7        | 0,2        |
| 8               | 0,4  | 0,5        | 0,2        | 0,6  | 0,5        | 0,2        |
| Média           | 0,4  | 0,4        | 0,1        | 0,6  | 0,6        | 0,2        |
| Desvio - Padrão | 0,0  | 0,1        | 0,1        | 0,1  | 0,1        | 0,1        |
| Máximo          | 0,4  | 0,5        | 0,2        | 0,7  | 0,7        | 0,3        |
| Mínimo          | 0,3  | 0,3        | 0,1        | 0,5  | 0,4        | 0          |

Fonte: Os autores (2021).

## 7 CONCLUSÃO

Para planejar uma rede GNSS devemos ter um conhecimento prévio aproximado da Matriz Variância-Covariância das linhas base. Pela dificuldade em se obter um valor prévio para as covariâncias, elas são adotadas como sendo zero. Aqui, testamos quais as implicações em se adotar covariâncias nulas e não nulas sobre os valores de confiabilidade da rede GNSS. As análises foram feitas sob a ótica da teoria da confiabilidade estendida apresentada em Rofatto (2020). Destacamos a seguir as principais conclusões decorrentes do trabalho:

- a) o valor crítico otimizado por MC é praticamente o mesmo para covariâncias nulas ou não nulas;
- b) embora os valores críticos sejam praticamente iguais, a confiabilidade melhora quando as covariâncias entre as componentes da linha base não são nulas (~30% para  $\Delta X$  e  $\Delta Y$ , e ~14% para  $\Delta Z$ );
- c) essa melhoria não depende do nível de significância ( $\alpha$ );
- d) para altos níveis de significância ( $\alpha > 0,1$ ) e sistemas com boa redundância ( $r_i > 0,5$ ), a confiabilidade para um modelo estocástico com covariâncias nulas se aproxima do caso em que as covariâncias não são nulas;
- e) na ausência de um modelo estocástico mais completo (covariâncias não-nulas) e para sistemas com boa redundância local ( $r_i > 0,5$ ), deve optar por maiores valores críticos para a estatística de teste do DS ( $\hat{k} < 2,8$ ).

Como recomendação para trabalhos futuros, sugere-se investigar cenários com covariâncias nulas e não nulas em redes com diferentes configurações geométricas e extensões de vetores, bem como para redes com diferentes coeficientes de correlação entre as componentes de cada linha-base.

### Agradecimentos

Essa pesquisa teve apoio financeiro da CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – proc. nº 88887.492336/2020-000). Agradecemos ao professor da Universidade Federal de Uberlândia, Prof. Dr. Gabriel do Nascimento Guimarães, e aos Eng. Agr. e Cartógrafos, Ramires Silva Rosa, Gabriel da Costa Resende e Valmir dos Reis Tavares pelo fornecimento dos dados RINEX referentes aos levantamentos do ano de 2016. Os comentários e sugestões dos revisores e editores melhoraram substancialmente o conteúdo do manuscrito. Por essa razão, os autores agradecem também os revisores e editores por suas contribuições. Por fim, agradecemos o importante apoio do grupo de pesquisa “Controle de Qualidade e Inteligência Computacional em Geodésia” ([dgp.cnpq.br/dgp/espelhogrupo/0178611310347329](http://dgp.cnpq.br/dgp/espelhogrupo/0178611310347329)).

### Contribuição dos Autores

Conceptualização e investigação: Rofatto, V. F., Bonimani, M. L. S., Matsuoka, M. T. e Klein, I.; Curadoria dos dados e visualização: Bonimani, M. L. S.; Análise formal e redação: Bonimani, M. L. S., Rofatto, V. F., Matsuoka, M. T.; Administração do Projeto e Supervisão: Rofatto, V. F., Matsuoka, M. T., Klein, I., Veronez, M. R.; Silveira Jr, L. G.; Software: Rofatto, V. F., Bonimani, M. L. S.; Validação: todos os autores; Revisão: todos os autores.

### Conflitos de Interesse

Os autores declaram que não há conflitos de interesse.

## Referências

- BAARDA, W. A. Testing Procedure for Use in Geodetic Networks. **Publications on Geodesy**, v. 2, n. 5, 97 p., 1968.
- COLLISCHONN, C.; MATSUOKA, M. T.; GUZZATTO, M. P.; KLEIN, I.; DE LIMA, E. M.; DE SOUZA, S. F. ANÁLISE DE CONFIABILIDADE DE REDE GNSS E REPETIÇÃO DE LINHAS-BASE. **Revista Brasileira de Cartografia**, v. 67, p. 1683-1703, 2015.
- DALLE, H. M. **Simulação do reator TRIGA IPR-RI utilizando métodos de transporte por Monte Carlo**. 179 f. Tese (Doutorado em Engenharia Química) – Universidade Estadual de Campinas, Campinas, 2005.
- FÖRSTNER, W. Reliability and discernability of extended Gauss-Markov models. In: SEMINAR ON MATHEMATICAL MODELS OF GEODETIC/PHOTOGRAMMETRIC POINT DETERMINATION WITH REGARD TO OUTLIERS AND SYSTEMATIC ERRORS, Germany: **Deutsche Geodätische Kommission**, 1983, p. 79–103.
- GHILANI, C. D. **Adjustment computations: spatial data analysis**. 5° ed. New Jersey: John Wiley & Sons, 2010.
- HAWKINS, D. M. **Identification of outliers**. London: Chapman & Hall, 1980.
- KLEIN, I.; MATSUOKA, M. T.; GUZZATTO, M. P.; NIEVINSKI, F. G. An approach to identify multiple outliers based on sequential likelihood ratio tests. **Survey Review**, v. 49, n. 357, p. 449-457, 2017. DOI. 10.1080/00396265.2016.1212970.
- KOCH, K. R. **Parameter estimation and hypothesis testing in linear models**. 2° ed. Berlin: Springer, 1999.
- LEHMANN, R. Improved critical values for extreme normalized and studentized residuals in Gauss–Markov models. **Journal of Geodetic Science**, v. 86, n. 12, p. 1137–1146, 2012. DOI. 10.1007/s00190-012-0569-0.
- LEHMANN, R.; VOß-BÖHME, A. On the statistical power of Baarda’s outlier test and some alternative. **Journal of Geodetic Science**, v. 7, n. 1, p. 68-78, 2017.
- LEICK, A.; RAPOPORT, L.; TATARNIKOV, D. **GPS Satellite Surveying**. 4° ed. John Wiley & Sons, 2015.
- MATSUMOTO, M.; NISHIMURA, T. Mersene Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. **ACM Transactions on Modeling and Computer Simulation**, v. 8, n. 1, 1998. DOI. 10.1145/272991.272995.
- MONICO, J. **Posicionamento pelo GNSS: descrição, fundamentos e aplicações**. 2° ed. Editora Unesp, 2008.
- NOWEL, K. Application of Monte Carlo method to statistical testing in deformation analysis based on robust M-estimation. **Survey Review**, v. 48, n. 348, p. 212-223, 2016.
- PRÓSZYŃSKI, W. Revisiting Baarda’s concept of minimal detectable bias with regard to outlier identifiability. **Journal of Geodesy**, v. 89, n. 10, p. 993-1003, 2015. DOI. 10.1007/s00190-015-0828-y.
- ROFATTO, V.; MATSUOKA, M.; KLEIN, I. An attempt to analyse Baarda’s iterative data snooping procedure based on Monte Carlo simulation. **South African Journal of Geomatics**, v. 6, n. 3, p. 416–435, 2017. DOI. 10.4314/sajg.v6i3.11.
- ROFATTO, V. F. **Uma extensão à teoria da confiabilidade em geodésia**. 2020. 255 f. Tese (Doutorado em Sensoriamento Remoto) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2020.
- ROFATTO, V. F.; MATSUOKA, M. T.; KLEIN, I.; VERONEZ, M. R.; BONIMANI, M. L.; LEHMANN, R. A half-century of Baarda’s concept of reliability: a review, new perspectives, and applications. **Survey Review**, v. 52, n. 372, p. 261-277, 2020a. DOI. 10.1080/00396265.2018.1548118.
- ROFATTO, V. F.; MATSUOKA, M. T.; KLEIN, I.; VERONEZ, M. R., DA SILVEIRA, L. G. A Monte Carlo Based outlier diagnosis method for sensitivity analysis. **Remote Sensing**, v. 12, n. 5, p. 1-41, 2020b. DOI. 10.3390/rs12050860.
- TEUNISSEN, P. **Testing Theory: an introduction**. 2° ed. Delft University Press, 2006.

YANG, L.; WANG, J.; KNIGHT, N. L.; SHEN, Y. Outlier separability analysis with a multiple alternative hypotheses test. **Journal of Geodesy Science**, v. 87, n. 6, p. 591–604, 2013. DOI.10.1007/s00190-013-0629-0.

### Biografia do autor principal



Maria Luísa Silva Bonimani, nasceu em Uberlândia, Minas Gerais, em 1997. Cursou Engenharia de Agrimensura e Cartográfica (2019) na Universidade Federal de Uberlândia. Atualmente é mestranda no Programa de Pós-Graduação em Agricultura e Informações Geoespaciais (2020-2022) da Universidade Federal de Uberlândia, campus Monte Carmelo. Durante a graduação desenvolveu pesquisas na área de Controle de Qualidade em Geodesia e no projeto de mestrado desenvolve pesquisas na área de Planejamento de Redes GNSS baseado em Simulação Computacional.



Esta obra está licenciada com uma Licença [Creative Commons Atribuição 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/) – CC BY. Esta licença permite que outros distribuam, remixem, adaptem e criem a partir do seu trabalho, mesmo para fins comerciais, desde que lhe atribuam o devido crédito pela criação original.