

Vinicius Francisco Rofatto

**Uma Extensão à Teoria da Confiabilidade
em Geodésia**

Porto Alegre - RS

2020

Vinicius Francisco Rofatto

Uma Extensão à Teoria da Confiabilidade em Geodésia

Tese apresentada ao Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia da Universidade Federal do Rio Grande do Sul para obtenção do título de Doutor em Sensoriamento Remoto pelo Programa de Pós-Graduação em Sensoriamento Remoto.
Linha de Pesquisa: Geodésia por Satélite.

Universidade Federal do Rio Grande do Sul – UFRGS

Centro Estadual de Pesquisa em Sensoriamento Remoto e Meteorologia

Programa de Pós-Graduação em Sensoriamento Remoto

Orientador: Prof. Dr. Marcelo Tomio Matsuoka

Coorientador: Prof. Dr. Ivandro Klein

Porto Alegre - RS

2020

CIP - Catalogação na Publicação

Rofatto, Vinicius Francisco
Uma Extensão à Teoria da Confiabilidade em Geodésia
/ Vinicius Francisco Rofatto. -- 2020.
234 f.
Orientador: Marcelo Tomio Matsuoka.

Coorientador: Ivandro Klein.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, Programa de Pós-Graduação em Sensoriamento Remoto, Porto Alegre, BR-RS, 2020.

1. Confiabilidade. 2. Outlier. 3. Teste de Hipóteses. 4. Controle de Qualidade. 5. Geodésia. I. Matsuoka, Marcelo Tomio, orient. II. Klein, Ivandro, coorient. III. Título.

PROGRAMA DE PÓS-GRADUAÇÃO EM SENSORIAMENTO REMOTO

TESE

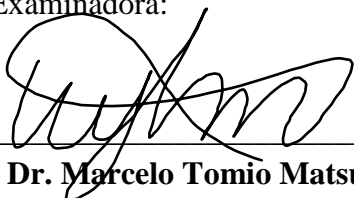
Submetida como parte dos requisitos
para obtenção do Grau de

DOCTOR EM SENSORIMENTO REMOTO

Programa de Pós-Graduação em Sensoriamento Remoto (PPGSR)
Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia (CEPRSM)
Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre, RS, Brasil.

Aprovada em: 11/05/2020

Pela Banca Examinadora:



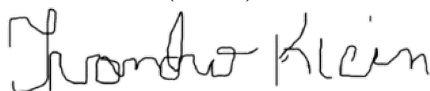
Prof. Dr. Marcelo Tomio Matsuoka

Orientador



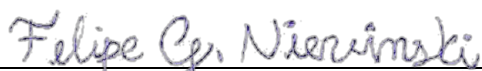
Profa. Dra. Eniuce Menezes Souza

(UEM)



Prof. Dr. Ivandro Klein

(IFSC)



Prof. Dr. Felipe Geremia Nievinski

(PPGSR/UFRGS)



Prof. Dr. Sérgio Florêncio de Souza

(PPGSR/UFRGS)

Homologada em: __/__/____

Por:

Prof. Dr. Guilherme Garcia de Oliveira

Coordenador do Programa de
Pós-Graduação em Sensoriamento
Remoto

Aos meus amados pais, Angela Cristina Barbosa Rofatto e Edson Francisco Rofatto.

A minha querida esposa, Alessandra Passalacqua Rofatto.

A minha querida filha, Bianca Passalacqua Rofatto

Agradecimentos

Inicio meus agradecimentos àqueles que me conduziram até o mundo da ciência. Aprendi que todo cientista é por natureza um pesquisador, mas nem todo pesquisador é um cientista. Sou grato ao meu orientador e amigo, Prof. Dr. Marcelo Tomio Matsuoka, uma pessoa curiosa, crítica com o que percebe e empreendedora para conseguir as respostas às suas indagações, um verdadeiro cientista por quem tenho admiração e respeito. Eu também não poderia produzir este trabalho sem os debates incansáveis com Prof. Dr. Ivandro Klein, do Instituto Federal de Santa Catarina (IFSC), um amigo por quem também tenho admiração e respeito. Cientistas são formados por cientistas.

Tanto a visão do Prof. Tomio quanto do Prof. Ivandro foram essenciais para o desenvolvimento deste trabalho. Quero também registrar aqui que, antes das publicações de outros cientistas renomados na área terem sido publicadas, a ideia de aplicar o Método Monte Carlo para descrever os níveis de probabilidade do procedimento *Data-Snooping* já estavam em nossas discussões. Embora informal, o registro consta em nossos debates no grupo WhatsApp, “DS Internal Reliability”. Sinto-me honrado por ter me tornado parte desse grupo de pesquisa e feliz por saber que novas propostas já vêm sendo planejadas em torno da aplicação de clássicas ferramentas estatísticas, porém adaptadas ao cenário atual do *Big Data problem*. Aliado a isso, o grupo também já vem propondo integração de modernas técnicas estatísticas para lidar com esse cenário, como mineração de dados (Data Mining) e aprendizado (ou aprendizagem) de máquina (Machine Learning).

Agradeço também os professores do Programa de Pós-Graduação em Sensorimento Remoto (PPGSR) pelas disciplinas ministradas. Os meus colegas do PPGSR pelo apoio nas atividades a distância. O Prof. Sérgio Florêncio de Souza e Prof. Felipe Nienvinski pelas contribuições no exame de qualificação. Os professores da Banca Examinadora. Os amigos da Universidade Federal de Uberlândia (UFU) por todo apoio oferecido durante a trajetória desta Tese em particular, ao Prof. Dr. João Vitor Meza Bravo pelas frequentes discussões sobre o que é uma Tese.

Agradeço imensamente a Coordenação e a Secretaria do PPGSR por todos os trâmites necessários para a realização desta Tese, principalmente pelo apoio nas atividades à distância.

Em seguida, gostaria de agradecer o Dr. Maurício Roberto Veronez e o Dr. Luiz Gonzaga da Silveira Junior do VizLab - Universidade do Vale do Rio dos Sinos (Unisinos) pelo reconhecimento e apoio financeiro para a publicação dos artigos desta Tese.

Um agradecimento muito especial aos meus pais, Edson e Angela, pelo amor e carinho que recebo durante todo o tempo. Ao meu irmão Felipe Augusto Rofatto e Marcelo

Prado pelas conversas sinceras e apoio ao longo desta jornada. Aos meus sogros, Sra. Leidemar e Sr. Braz, que durante o tempo desta Tese me apoiaram e acreditaram no meu trabalho. Um muito obrigado aos membros da minha família, Tio Edvaldo e Maria Silvia, meus primos Henrique e Layla, Otávio e Tamara, e Vitor. Em especial, dedico esta Tese aos meus avós Zela e Chico (*in memoriam*).

Como sempre, deixamos a cereja do bolo para o final. Gostaria de agradecer a minha esposa Alessandra, que durante esta fase tornou-se uma grande mãe e teve muito carinho, paciência e compreensão para comigo, principalmente nos tempos de estresse. À minha filha, Bianca, pelo amor e carinho.

Obrigado a todos!

*“Não vos amoldeis às estruturas deste mundo,
mas transformai-vos pela renovação da mente,
a fim de distinguir qual é a vontade de Deus:
o que é bom, o que Lhe é agradável, o que é perfeito.
(Bíblia Sagrada, Romanos 12, 2)*

Resumo

Há mais de meio século, a teoria da confiabilidade introduzida por Baarda (1968) tem sido usada como uma prática padrão para o controle de qualidade em geodésia. Embora atenda o rigor matemático e os pressupostos da probabilidade, a teoria foi originalmente desenvolvida para um Data-Snooping que considera uma específica observação como sendo um outlier. Na prática, não sabemos qual observação é um outlier. Se o objetivo do procedimento Data-Snooping é testar cada observação individual quanto à presença de um outlier, então uma hipótese alternativa mais apropriada seria: “Existe pelo menos um outlier nos dados observados”. Agora, estamos interessados em responder: “Onde?”. A resposta para tal pergunta recai sobre um problema de localizar dentre as hipóteses alternativas aquela que levou à rejeição da hipótese nula, ou seja, estamos interessados em identificar o outlier. Esse problema é conhecido como múltiplas hipóteses alternativas. Embora avanços tenham ocorrido ao longo desse período, as teorias apresentadas até o momento consideram apenas uma única rodada do Data-Snooping, sem qualquer diagnóstico subsequente, como a remoção do outlier. Na prática, entretanto, o Data-Snooping é aplicado de forma iterativa: após a identificação e a eliminação de um possível outlier, os dados são reprocessados e a identificação é reiniciada. Este procedimento é denominado de Data-Snooping Iterativo (DSI). O DSI é, portanto, um caso que envolve não somente múltiplas hipóteses alternativas, mas também múltiplas rodadas de estimação, teste e adaptação. Estimar os níveis de probabilidade associado com DSI é praticamente impossível por aqueles métodos analíticos usualmente empregados em procedimentos mais simples, por exemplo, o teste global do modelo e Data-Snooping de uma única hipótese alternativa. Por essa razão, uma rigorosa e completa teoria da confiabilidade não estava disponível até o momento. Embora grandes avanços tenham ocorrido em meados da década de 1970, como os computadores baseados em microprocessadores, Baarda tinha uma desvantagem: a tecnologia de sua época era insuficiente para que se utilizassem técnicas computacionais inteligentes. Hoje o cenário computacional é completamente diferente da época da teoria da confiabilidade de Baarda. Aqui, seguindo a tendência atual da ciência moderna, usamos o método de Monte Carlo e estendemos a teoria da confiabilidade para o DSI. Neste trabalho, demonstramos que a estimação depende do teste e da adaptação e, portanto, o DSI é, na verdade, um estimador. Até o presente momento, a escolha do número de simulações de Monte Carlo tem sido avaliada somente em função da precisão. Assim, levantou-se uma questão: como podemos encontrar um número ótimo de experimentos Monte Carlo em termos de acurácia? Aqui, usamos eventos com probabilidades conhecidas para avaliar a acurácia do Método de Monte Carlo. Os resultados mostraram que, dentre os números de experimentos testados, $m = 200,000$ forneceu suficiente precisão numérica, com erro relativo menor que 0.1%. A estatística de teste associada ao DSI é o valor extremo dos resíduos dos mínimos quadrados normalizados. É bem conhecido na literatura que valores críticos desse teste não podem

ser derivados de distribuições conhecidas, mas devem ser calculados numericamente por meio do método de Monte Carlo. Este trabalho fornece os primeiros resultados sobre o valor crítico baseado em Monte Carlo inserido em diferentes cenários de correlação entre as estatísticas de teste. Testamos se o aumento do nível de significância conjunto, ou redução do valor crítico, melhora a identificabilidade do outlier. Os resultados mostraram que quanto menor o valor crítico, ou maior o nível de significância conjunto, maior é a probabilidade de correta detecção, e menor é o MDB. Porém, essa relação não é válida em termos de identificação. Observamos que, quando o efeito de todas as observações na taxa de falsa exclusão (Erro Tipo III) diminui, é possível encontrar o menor outlier identificável (MIB). A razão disso é que o efeito da correlação entre os resíduos torna-se insignificante para uma certa magnitude de outlier, o que aumenta a probabilidade da correta identificação.

Palavras-chave: Probabilidade. Testes de Hipóteses. Detecção de Outlier. Simulação Estocástica. Gerador de Números Aleatórios. Monte Carlo. Controle de Qualidade. Confiabilidade.

Abstract

For more than half a century, the reliability theory introduced by [Baarda \(1968\)](#) has been used as a standard practice for quality control in geodesy and surveying. Although the theory meets mathematical rigor and probability assumptions, it was originally developed for a Data-Snooping which assumes a specific observation as a suspect outlier. In other words, only one single alternative hypothesis is in play. Actually, we do not know which observation is an outlier. Since the Data-Snooping consists of screening each individual measurement for an outlier, a more appropriate alternative hypothesis would be: “There is at least one outlier in the observations”. Now, we are interested to answer: “Where?”. The answer to this question lies in a problem of locating among the alternative hypotheses the one that led to the rejection of the null hypothesis. Therefore, we are interested in identifying the outlier. Although advances have occurred over that period, the theories presented so far consider only one single round of the Data-Snooping procedure, without any subsequent diagnosis, such as removing the outlier. In fact, however, Data-Snooping is applied iteratively: after identification and elimination of the outlier, the model is reprocessed, and outlier identification is restarted. This procedure of iterative outlier elimination is known as Iterative Data-Snooping (IDS). Computing the probability levels associated with IDS is virtually impossible to those analytical methods usually employed in conventional tests, such as, overall model test and Data-Snooping of only one single alternative hypothesis. Because of this, a rigorous and complete reliability theory was not yet available. Although major advances occurred in the mid-1970s, such as microprocessor-based computers, Baarda had a disadvantage: the technology of his time was insufficient to use intelligent computational techniques. Today, the computational scenario is completely different from the time of Baarda’s theory of reliability. Here, following the current trend of modern science, we can use intelligent computing and extend the reliability theory when the DSI is in play. We show that the estimation depends on the test and the adaptation and, therefore, the IDS is, in fact, an estimator. Until the present, no study has been conducted to evaluate empirically the accuracy of the Monte Carlo for quality control purposes in geodesy. Generally, only the degree of dispersion of the Monte Carlo is considered. Thus, an issue remains: how can we find the optimal number of Monte Carlo experiments for quality control purpose? Here, we use an exact theoretical reference probabilities to answer this question. We find that that the number of experiments $m = 200,000$ can provide consistent results with sufficient numerical precision for outlier identification, with a relative error less than 0.1%. The test statistic associated with IDS is the extreme normalised least-squares residual. It is well-known in the literature that critical values (quantile values) of such a test statistic cannot be derived from well-known test distributions but must be computed numerically by means of Monte Carlo. This paper provides the first results on the Monte Carlo-based critical value inserted into different scenarios of correlation between outlier

statistics. We also tested whether increasing the level of the family-wise error rate, or reducing the critical values, improves the identifiability of the outlier. The results showed that the lower critical value, or the higher the family-wise error rate, the larger the probability of correct detection, and the smaller the MDB. However, this relationship is not valid in terms of identification. We also highlight that an outlier becomes identifiable when the contributions of the observations to the wrong exclusion rate (Type III error) decline simultaneously. In this case, we verify that the effect of the correlation between outlier statistics on the wrong exclusion rate becomes insignificant for a certain outlier magnitude, which increases the probability of identification.

Keywords: Probability. Hypothesis Testing. Outlier Detection. Stochastic Simulation. Random Number Generator. Monte Carlo. Quality Control. Reliability.

Lista de ilustrações

Figura 1 – Exemplo de aplicação do <i>Data-Snooping</i> em um modelo de regressão linear.	28
Figura 2 – Resultado do rompimento das barragens em Brumadinho/MG (a) e Mariana/MG (b).	31
Figura 3 – Exemplo da relação entre o parâmetro de não-centralidade e os níveis de probabilidade associados com o Data-Snooping.	38
Figura 4 – Efeito do aumento do número de testes sobre (α')	41
Figura 5 – Valores críticos calculados pela relação de Bonferroni.	42
Figura 6 – Decomposição dos efeitos de um outlier no espaço das observações, parâmetros e resíduos.	49
Figura 7 – Resultado de $\hat{\mathbf{x}}$ em função do procedimento de teste.	51
Figura 8 – Ilustração da aplicação do Método de Monte Carlo na propagação de distribuições de probabilidade.	57
Figura 9 – Exemplo de aplicação do Método Monte Carlo na descrição de modelos sob diferentes hipóteses.	57
Figura 10 – Fluxograma para estimação das probabilidades do DSI e determinação dos indicadores de sensibilidade (MDB e MIB).	62
Figura 11 – Relação entre Precisão, Veracidade e Acurácia.	67
Figura 12 – Relação entre a contribuição de cada observação identificada erroneamente para a falsa exclusão (\mathcal{P}_{WE}) e a taxa de correta identificação (\mathcal{P}_{CI}).	121
Figura 13 – Configurações da rede geodésica sujeita às diferentes injunções: altitude desconhecida (unknown heights) representado por círculo na cor azul, injunção absoluta (hard constraints) por triângulo em vermelho e injunção relativa (soft constraints) por círculo em vermelho.	164
Figura 14 – Probabilidades de correta identificação (\mathcal{P}_{CI}) e detecção (\mathcal{P}_{CD}) para $\hat{k} = 3.89$, no caso de uma única injunção absoluta, e $\hat{k} = 3.93$ para o caso de duas e três injunções absoluta. Cluster 1 (a,b): y_1, y_3, y_4 e y_6 . Cluster 2 (c,d): y_2 e y_5 . Cluster 3 (e,f): y_7, y_8, y_9 e y_{10} . Cluster 4 (g,h): y_{11} e y_{12}	166
Figura 15 – Probabilidades de falsa exclusão (\mathcal{P}_{WE}) para $\hat{k} = 3.89$ no caso de uma única injunção absoluta e $\hat{k} = 3.93$ para o caso de duas e três injunções absoluta. Cluster 1 (a): y_1, y_3, y_4 e y_6 . Cluster 2 (b): y_2 e y_5 . Cluster 3 (c): y_7, y_8, y_9 e y_{10} . Cluster 4 (d): y_{11} e y_{12}	167
Figura 16 – Taxa de sobreposição estatística (\mathcal{P}_{ol}) e probabilidade de correta detecção (\mathcal{P}_{CD}) do Cluster 1 para o cenário de uma única injunção absoluta.	168

Figura 17 – Probabilidades de correta detecção (\mathcal{P}_{CD}), identificação (\mathcal{P}_{CI}), falsa exclusão (\mathcal{P}_{WE}) e sobre-identificação (\mathcal{P}_{over+}) para $\alpha' = 0.001$ e $\alpha' = 0.1$. Resultados para o Cluster 1 (y_1, y_3, y_4 e y_6) quando sujeito à duas e três injunções absolutas.	169
Figura 18 – Cenário para duas injunções relativas. Probabilidades de correta identificação (\mathcal{P}_{CI}) e detecção (\mathcal{P}_{CD}) para $\hat{k} = 3.95$, no caso de $\sigma_c = 0.1mm$ e $\sigma_c = 1mm$; e $\hat{k} = 3.92$ para o caso de $\sigma_c = 10mm$. Cluster 1 (a,b): y_1, y_3, y_4 e y_6 . Cluster 2 (c,d): y_2 e y_5 . Cluster 3 (e,f): y_7, y_8, y_9 e y_{10} . Cluster 4 (g,h): y_{11} e y_{12}	171
Figura 19 – Cenário para duas injunções relativas. Probabilidade de correta detecção (\mathcal{P}_{CD}) para $\hat{k} = 3.92$, no caso de $\sigma_c = 0.1mm$ e $\sigma_c = 1mm$; e $\hat{k} = 3.92$ para o caso de $\sigma_c = 10mm$. Cluster 5: y_{13} e y_{14}	172
Figura 20 – Cenário para duas injunções relativas. Probabilidade de falsa exclusão (\mathcal{P}_{WE}) para $\hat{k} = 3.95$, no caso de $\sigma_c = 0.1mm$ e $\sigma_c = 1mm$; e $\hat{k} = 3.92$ para o caso de $\sigma_c = 10mm$. Cluster 1 (a): y_1, y_3, y_4 e y_6 . Cluster 2 (b): y_2 e y_5 . Cluster 3 (c): y_7, y_8, y_9 e y_{10} . Cluster 4 (d): y_{11} e y_{12}	172
Figura 21 – Cenário para três injunções relativas. Probabilidades de correta identificação (\mathcal{P}_{CI}) e detecção (\mathcal{P}_{CD}) para $\hat{k} = 3.99$, no caso de $\sigma_c = 0.1mm$ e $\sigma_c = 1mm$; e $\hat{k} = 3.96$ para o caso de $\sigma_c = 10mm$. Cluster 1 (a,b): y_1, y_3, y_4 e y_6 . Cluster 2 (c,d): y_2 e y_5 . Cluster 3 (e,f): y_7, y_8, y_9 e y_{10} . Cluster 4 (g,h): y_{11} e y_{12}	174
Figura 22 – Cenário para três injunções relativas. Probabilidades de correta identificação (\mathcal{P}_{CI}) e detecção (\mathcal{P}_{CD}) para $\hat{k} = 3.99$, no caso de $\sigma_c = 0.1mm$ e $\sigma_c = 1mm$; e $\hat{k} = 3.96$ para o caso de $\sigma_c = 10mm$. Cluster 5 (a,b): y_{13} e y_{14} . Cluster 6 (c,d): y_{15}	175
Figura 23 – Cenário para três injunções relativas. Probabilidade de falsa exclusão (\mathcal{P}_{WE}) para $\hat{k} = 3.99$, no caso de $\sigma_c = 0.1mm$ e $\sigma_c = 1mm$; e $\hat{k} = 3.96$ para o caso de $\sigma_c = 10mm$. Cluster 1 (a,b): y_1, y_3, y_4 e y_6 . Cluster 2 (c,d): y_2 e y_5 . Cluster 3 (e,f): y_7, y_8, y_9 e y_{10} . Cluster 4 (g,h): y_{11} e y_{12}	175

Lista de quadros

Quadro 1 – Sumário da evolução da teoria da confiabilidade.	69
---------------------------------------------------------------------	----

Lista de tabelas

Tabela 1 – Níveis de probabilidade associado com o <i>Data-snooping</i> sob múltiplas hipóteses alternativas	44
Tabela 2 – Redundância local (r_i), desvio-padrão do outlier ($\sigma_{\hat{v}_i}$) e máxima correlação absoluta entre os resíduos ($max_{\rho_{w_i, w_j}}$) para cada cenário de restrição rígida.	167
Tabela 3 – Redundância local (r_i), desvio-padrão do outlier ($\sigma_{\hat{v}_i}$) e máxima correlação absoluta entre os resíduos ($max_{\rho_{w_i, w_j}}$) para cada cenário de duas injunções relativas.	170
Tabela 4 – Redundância local (r_i), desvio-padrão do outlier ($\sigma_{\hat{v}_i}$) e máxima correlação absoluta entre os resíduos ($max_{\rho_{w_i, w_j}}$) para cada cenário de três injunções relativas.	173
Tabela 5 – MDB e MIB para o caso de injunções absolutas com $\alpha' = 0.001$ e $\tilde{\mathcal{P}}_{CD} = \tilde{\mathcal{P}}_{CI} = 0.8$	176
Tabela 6 – MDB and MIB para o caso de duas injunções relativas com $\alpha' = 0.001$ e $\tilde{\mathcal{P}}_{CD} = \tilde{\mathcal{P}}_{CI} = 0.8$	176
Tabela 7 – MDB and MIB para o caso de três injunções relativas com $\alpha' = 0.001$ e $\tilde{\mathcal{P}}_{CD} = \tilde{\mathcal{P}}_{CI} = 0.8$	176

Lista de abreviaturas e siglas

BLUE	<i>Best Linear Unbiased Estimator</i> (Melhor Estimador Linear Imparcial)
car	característica (ou posto) de uma matriz
det	determinante de uma matriz
DIA	Detecção, Identificação e Adaptação
DS	Data-Snooping
DSI	Data-Snooping Iterativo
ECDF	<i>Empirical Cumulative Density Function</i>
f.d.p.	função densidade de probabilidade
f.d.a.e.	função de distribuição acumulada empírica
FDR	<i>False Discovery Rate</i>
FWER	Family-Wise Error Rate (Taxa de Erro da Família dos Testes)
GLR	<i>Generalized Likelihood Ratio</i> (Teste da Razão de Verossimilhança Generalizado)
GNA	Gerador de Números Aleatórios
GNPA	Geradores de Números Pseudo-Aleatórios
GNSS	<i>Global Navigation Satellite System</i> (Sistema de Navegação Global por Satélite)
h.c.	injunção absoluta (<i>hard constraints</i>)
HRNG	<i>Hardware-Random Number Generator</i>
MDB	<i>Minimal Detectable Bias</i>
MIB	<i>Minimal Identifiable Bias</i>
MLE	<i>Maximum Likelihood Estimation</i> (Estimador de Máxima Verossimilhança)
MMC	Método de Monte Carlo
MMQ	Método dos Mínimos Quadrados

MSE	<i>Mean Squared Error</i>
PRNG	<i>Pseudo-Random Number Generator</i>
s.c.	injunção relativa (<i>soft constraints</i>)
SLR	<i>Simple Likelihood Ratio</i> (Teste da Razão de Verossimilhança Simples)
SLRTMO	<i>Sequential Likelihood Ratio Tests for Multiple Outliers</i>
UMP	<i>Uniformly Most Powerful</i> (Teste Uniformemente Mais Poderoso)
WLSE	<i>Weighted Least Squares Estimation</i> (Estimador de Mínimos Quadrados Ponderados)

Lista de símbolos

α_0	nível de significância
α_i	nível de significância de cada teste individual
α'	nível de significância conjunto
β_0	probabilidade de cometer o Erro Tipo II para um teste de uma única hipótese alternativa
β_{i0}	probabilidade de cometer o Erro Tipo II para o caso de múltiplas hipóteses alternativas
β_{i0}	probabilidade de cometer o Erro Tipo II para o caso de múltiplas hipóteses alternativas
β_{ii}	probabilidade de cometer o Erro Tipo II e Erro Tipo III para o caso de múltiplas hipóteses alternativas
γ_0	poder do teste para o caso de uma única hipótese alternativa
κ_{ij}	probabilidade de cometer o Erro Tipo III
λ_0	parâmetro de não-centralidade para um teste de uma única hipótese alternativa
$\lambda_{q=1}^{(MDB)}$	parâmetro de não-centralidade associado ao menor outlier detectável
$\lambda_{q=1}^{(MIB)}$	parâmetro de não-centralidade associado ao menor outlier identificável
ε	vetor de incertezas decomposto em dois componentes: um componente associado apenas aos efeitos aleatórios e outro componente associado com o outlier
$\hat{\varepsilon}$	ε estimado
ρ_{w_i, w_j}	coeficiente de correlação entre as estatísticas de teste
$\max_{\rho_{w_i, w_j}}$	máximo valor absoluto do coeficiente de correlação entre as estatísticas de teste
$\sigma_{\hat{\tau}}$	desvio-padrão da estimativa do outlier
σ_c	desvio-padrão da injunção

χ^2	distribuição qui-quadrado
∇	vetor de parâmetros incógnitos associado à magnitude dos outliers
∇_i	escalar que corresponde à magnitude do outlier
∇_{wi}	escalar que corresponde à raiz quadrada do parâmetro de não-centralidade da distribuição χ^2
∇_y	efeito do outlier nas observações
$\nabla_{\hat{x}}$	efeito do outlier nos parâmetros
$\nabla_{\hat{\epsilon}}$	efeito do outlier nos resíduos
P_A	projektor que projeta sobre o espaço coluna da matriz A e na direção de seu complemento ortogonal
\perp	complemento ortogonal
A_{α_0}	região de aceitação para um teste de uma única hipótese alternativa
A_0	região de aceitação para o DSI
A_i	região crítica para o DSI
\mathcal{F}	função
\mathcal{H}_0	hipótese nula
\mathcal{H}_A	hipótese alternativa
$\mathcal{H}_A^{(i)}$	hipóteses alternativas
\mathcal{N}	função de densidade de probabilidade normal
\mathcal{P}	probabilidade
\mathcal{R}	espaço coluna de uma matriz
\mathcal{R}^\perp	complemento ortogonal do espaço coluna de uma matriz
\mathbb{D}	operador de variância
\mathbb{E}	operador de esperança matemática
\mathbb{R}	espaço dos números reais
n	número de observações
u	número de parâmetros

q	número de outliers ou graus de liberdade do teste
gl	graus de liberdade local
r	redundância global do modelo
r_i	número de redundância local
k_0	valor crítico para um teste de uma única hipótese alternativa
k_i	valor crítico para cada teste individual
k_{bonf}	valor crítico calculado a partir da relação de Bonferroni
\hat{k}	valor crítico obtido empiricamente via método de Monte Carlo
F_N^{-1}	inversa da função de distribuição acumulada da normal, ou também chamada de função quantil da distribuição normal
A	matriz <i>design</i> (ou jacobiana)
C	matriz que define o modelo de outlier
R	matriz redundância
c_i	vetor que define o modelo de outlier
e	vetor das incertezas
\hat{e}	vetor das incertezas estimadas ou vetor dos resíduos estimados
$Q_{\hat{e}}$	matriz covariância dos resíduos estimados
Q_e	matriz covariância das observações
W	matriz peso das observações
$Q_{\hat{y}}$	matriz covariância das observações estimadas
y	vetor das observações (ou observáveis)
\hat{y}	vetor das observações (ou observáveis) estimadas
x	vetor dos parâmetros incógnitos do modelo
\hat{x}	vetor dos parâmetros estimados
w_i	estatística de teste associado ao DS
T_q	teste de razão de verossimilhança generalizado

$\text{max-}w$	máximo valor absoluto da estatística de teste w_i
\tilde{w}	sequências de $\text{max-}w$ ordenados em ordem ascendente
\mathcal{R}_w	matriz de correlação entre as estatísticas de teste w_i
$f_y(\mathbf{y} \mathbf{x})$	função densidade de probabilidade de um vetor aleatório \mathbf{y}

Sumário

1	INTRODUÇÃO	24
1.1	Justificativa	29
1.2	Objetivos Geral e Específicos	31
1.3	Estrutura da Tese	31
2	FUNDAMENTAÇÃO TEÓRICA	35
2.1	Teoria da Confiabilidade Convencional	35
2.2	Teoria da Confiabilidade no contexto de Múltiplas Hipóteses Alternativas	40
2.3	Uma Extensão à Teoria da Confiabilidade	46
2.4	Método de Monte Carlo no contexto da Geodésia	54
3	MATERIAL E MÉTODOS	59
3.1	Procedimento para obtenção do valor crítico da estatística <i>max-w</i>	60
3.2	Procedimento para estimação dos níveis de probabilidade do Data-Snooping Iterativo	61
4	EXPERIMENTOS E RESULTADOS	66
4.1	Avaliação da acurácia do Método de Monte Carlo	67
4.2	Avaliação do valor crítico obtido por meio do Método de Monte Carlo	117
4.3	Análise do efeito do valor crítico controlado sobre os níveis de probabilidade do Data-Snooping	118
4.4	Estudo da relação entre as medidas analíticas de controlabilidade e os níveis de probabilidade do Data-Snooping	120
4.5	Avaliação dos efeitos das injunções absolutas e relativas sobre a sensibilidade do Data-Snooping	163
5	CONSIDERAÇÕES FINAIS	197
5.1	Principais Conclusões	198
5.2	Perspectivas Futuras	199
	REFERÊNCIAS	201
	APÊNDICES	208
	APÊNDICE A – TEORIA DA ESTIMAÇÃO	209

A.1	Propriedades ótimas dos estimadores	210
A.2	Método dos mínimos quadrados ponderados	212
A.3	Melhor estimador linear imparcial	215
A.4	Estimativa de máxima verossimilhança	216
	APÊNDICE B – TESTES DE HIPÓTESES EM MODELOS LINE-	
	ARES	218

1 Introdução

Há mais de meio século, o *Data-Snooping*, inicialmente proposto por Baarda (1968), tem sido usado como um procedimento padrão para a detecção de outliers em geodésia (KLEIN, 2011; LEHMANN, 2012; LEHMANN; VOß-BÖHME, 2017; KOCH, 1999; GHILANI, 2017; LEICK; RAPOPORT; TATARNIKOV, 2015; ROFATTO et al., 2020a). Embora originalmente introduzido para o controle de qualidade de redes geodésicas¹, o *Data-Snooping* não se restringe ao campo da geodésia, mas é um método que pode ser aplicado em qualquer área do conhecimento (ROFATTO et al., 2020a).

A definição de outlier mais frequentemente encontrada na literatura é dada por Hawkins (1980): “um outlier é uma observação que se diferencia tanto das demais observações que levanta suspeitas de que foi gerada por um mecanismo distinto”.

Geralmente, outliers e erros grosseiros são tratados como sinônimos. A definição de erro grosseiro (*blunder*) pode ser dada como: “erros grosseiros são causados por erros humanos, instrumentos com defeito ou métodos incorretos de medição. Erros grosseiros não seguem certas regras e normalmente não podem ser tratados no ajustamento por mínimos quadrados clássico. Em princípio, erros grosseiros não são permitidos e devem ser evitados” (GEMAEL, 1994).

Em geodésia, os outliers são frequentemente causados por erros grosseiros e erros grosseiros geralmente causam outliers (LEHMANN, 2013b). Por essa razão, muitas vezes eles são confundidos e empregados erroneamente. Na verdade, outliers podem ser resultados de medições completamente corretas, e erros grosseiros nem sempre levam a grandes desvios. Em termos de erro grosseiro, pode-se registrar uma medida com o valor 3.57² da unidade, quando na verdade deveria ter sido 3.75. Embora tenha cometido um erro grosseiro, a magnitude do erro de 0.18 pode estar dentro do intervalo de incerteza do modelo estipulado e, portanto, o valor 3.57 não será outlier. Por outro lado, a natureza, por si só, às vezes é responsável por gerar anomalias nas observações. Por exemplo, hoje a fonte mais preocupante de anomalias no GNSS é a ionosfera (MATSUOKA, 2007). Se esse efeito puder ser quantificado e, se for significativo com relação à exatidão requerida da medição, uma correção ou parametrização no modelo matemático poderão ser aplicados para compensar o efeito. Após esta correção, ou parametrização, a esperança da distribuição de probabilidade

¹ Rede geodésica: conjunto de pontos devidamente materializados no terreno, cujas coordenadas (posições) em relação a um referencial são estimadas por meio de observações terrestres como medidas de direções e distâncias entre os vértices, e/ou por meio de técnicas espaciais como os métodos de posicionamento por satélites GNSS (*Global Navigation Satellite System*) (KLEIN, 2011). Esse conjunto de pontos, estruturalmente estáveis e identificáveis, são interconectados por meio de medidas de direção, distância ou diferença de nível, formando uma rede geodésica.

² NOTA: separador decimal utilizado ao longo deste trabalho é “ponto”.

associado ao efeito sistemático será zero.

Com base nessas premissas, [Lehmann \(2013b\)](#) propõe a seguinte definição: “outlier é uma observação que provavelmente é causada por um erro grosseiro que é melhor ser descartado ou não ser usado como observação”.

Embora as definições acima sejam aceitáveis, ainda assim, elas não são completas. No caso da definição de Hawkins, não está claro se o desvio da observação é com relação a um modelo ou padrão conhecido. No caso da definição de Lehmann, a causa de um outlier está associada com a ocorrência do erro grosseiro. Com base nessas discussões, aqui propomos a seguinte definição: **“outlier é uma observação que se afastou de seu valor mais provável a ponto de não pertencer ao modelo matemático (funcional e estocástico) estipulado”**. Para garantir, portanto, as premissas e pressupostos do modelo matemático estipulado de um dado problema, outliers devem ser adequadamente tratados.

Importante destacar que os modelos geralmente são configurados de uma maneira que seus parâmetros atendam certas restrições. Tais restrições se originam de um conhecimento prévio do cientista. Dentro do contexto de geodésia, por exemplo, é comum adotar as injunções absolutas e relativas. As injunções absolutas geralmente são adotadas nos casos em que existem relações funcionais exatamente conhecidas entre os parâmetros desconhecidos. Nesse caso, as injunções são tratadas como constantes e não como variáveis aleatórias. Por outro lado, as injunções relativas são tratadas como variáveis aleatórias e, portanto, podem violar tais relações funcionais em função de suas incertezas. Esses tipos de restrições também são alvos de investigações nesta tese.

Duas categorias de técnicas avançadas para o tratamento de um conjunto de dados contaminado por outliers têm sido frequentemente desenvolvidas e aplicadas em várias situações: estimação robusta ([ROUSSEEUW](#); [LEROY, 2003](#); [YANG, 1999](#); [WILCOX, 2013](#); [DUCHNOWSKI, 2013](#); [KLEIN et al., 2015](#); [SURACI](#); [OLIVEIRA, 2010](#)) e procedimentos baseados em testes de hipóteses ([BAARDA, 1967](#); [BAARDA, 1968](#); [ARNOLD, 1981](#); [HAWKINS, 1980](#); [FÖRSTNER, 1983](#); [YANG et al., 2013](#); [ROFATTO](#); [MATSUOKA](#); [KLEIN, 2017](#)). No primeiro caso, os efeitos dos outliers nos parâmetros incógnitos do modelo são minimizados por meio de um estimador dito ser robusto. Em outras palavras, a estimação robusta consiste em minimizar a influência de possíveis outliers sobre os parâmetros do modelo. As classes dessa técnica incluem M-estimadores (“M” de máxima verossimilhança), cuja estatística é baseada no método de máxima verossimilhança, veja por exemplo [Huber \(1981\)](#); L-estimadores (“L” de linear) são derivados da combinação linear de estatísticas de ordem (máximo, mínimo, mediana e outros quantis amostrais); e o R-estimador (“R” de *rank*) que envolve a classificação dos resíduos. Embora existam vantagens indubitáveis do método de estimação robusta, aqui, o foco é dado para o procedimento baseado em testes de hipóteses. Detalhes sobre testes de hipóteses podem

ser vistos no Apêndice (B).

As seguintes vantagens de testar o modelo quanto à possíveis outliers são elencadas a seguir (LEHMANN, 2013a):

1. Oportunidade de investigar as causas dos outliers.
2. Outliers identificados podem ser substituídos por novas medidas no modelo.
3. Com o objetivo de garantir a confiabilidade nos resultados de um sistema que opera sob o método dos mínimos quadrados (MMQ), o outlier identificado pode ser eliminado dos dados observados.

O último item acima (3) está diretamente associado com o procedimento estatístico *Data-Snooping*. A maioria dos softwares de processamento de dados geodésicos operam sob o método dos mínimos quadrados (MMQ), seja em tempo real, seja no modo pós-processado. Supondo que não exista outlier nos dados observados, o MMQ é o melhor estimador linear imparcial (*best linear unbiased estimator* – BLUE) (TEUNISSEN, 2006). Para detalhes sobre as propriedades dos estimadores, consultar o Apêndice (A.1). No entanto, os outliers podem inevitavelmente ocorrer na prática e causar a perda da propriedade BLUE do MMQ. Com a intenção de garantir as propriedades do MMQ³, o *Data-Snooping* tem a tarefa de identificar possíveis outliers. Por esta razão, o *Data-Snooping* também se tornou muito popular e é rotineiramente usado como procedimento padrão no controle de qualidade do ajustamento de observações geodésicas (KLEIN, 2011; GHILANI, 2017).

A Figura (1) ilustra um exemplo de aplicação do *Data-Snooping* em um problema de regressão linear com $n = 8$ observações. O eixo vertical dos gráficos são os resíduos dos mínimos quadrados normalizados.

Modelos de regressão é um padrão encontrado em qualquer área do conhecimento. Alguns exemplos em geodésia são (LEHMANN, 2015):

- extrair uma tendência linear de uma série temporal em estudos geodésicos ou geofísicos;
- ajustar uma função para fins de calibração de instrumentos de medição;
- ajustar uma linha reta espacial, cujos pontos se desviam do modelo da reta devido às incertezas associadas às medições topográficas.

³ A combinação de estimação e teste estatístico sempre leva à um estimador tendencioso, exceto o caso em que a hipótese nula é verdadeira (TEUNISSEN, 2018).

Neste caso, o modelo sob a hipótese nula, denotado por \mathcal{H}_0 , representa o modelo sem a presença de outlier. Ele foi formulado da seguinte forma:

$$\mathcal{H}_0 : y_i = ax_i + b + e_i, \forall i = 1, \dots, n, \text{ com } e_i \sim N(0, 1) \quad (1.1)$$

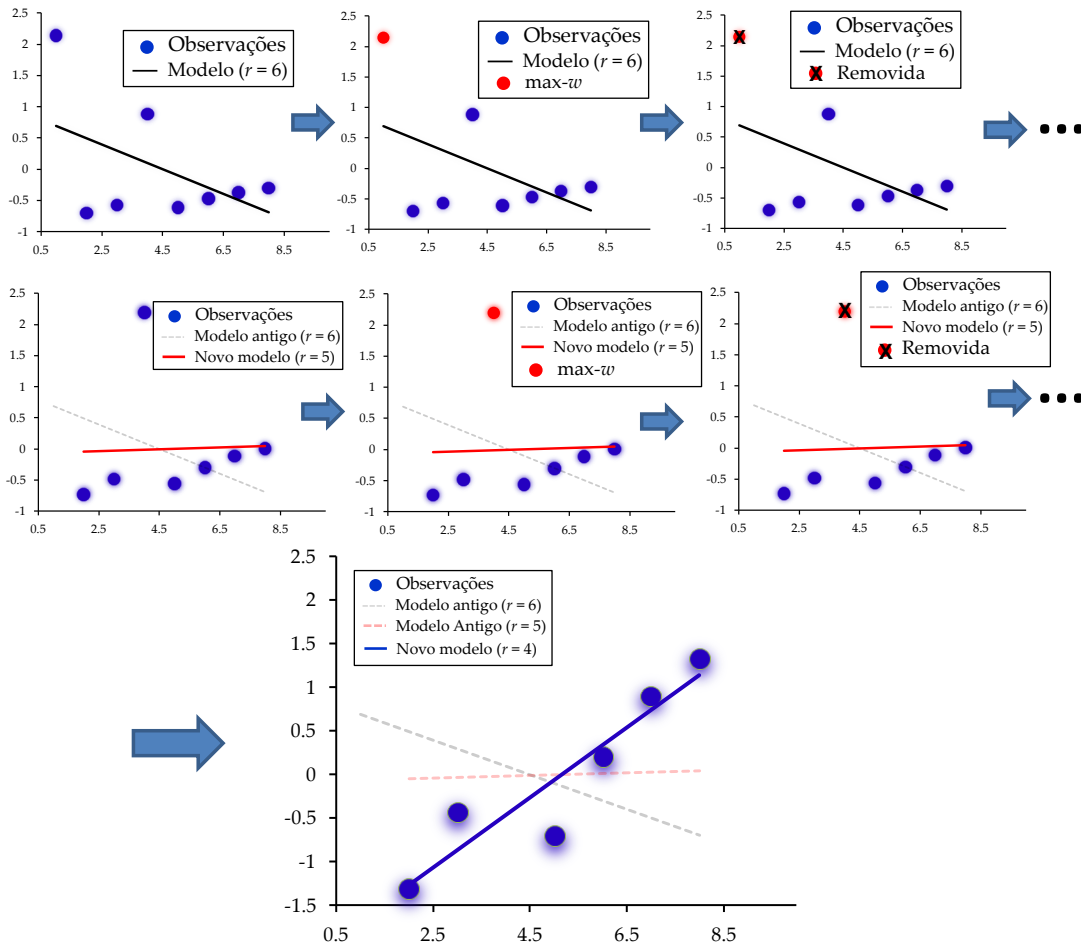
em que $y \in \mathbb{R}^{n \times 1}$ é o vetor contendo os dados observados, a e $b \in \mathbb{R}^{u \times 1}$ são os parâmetros incógnitos do modelo e $e_i \in \mathbb{R}^{n \times 1}$ descreve a aleatoriedade de y devido às inevitáveis fontes de incerteza (i.e. é um vetor de incertezas desconhecido). Observe que y depende de x , mas x não depende de y . Por esta razão, y é chamado de variável dependente e x a variável independente (ou variável explanatória). Aqui, os valores de x são fixos, isto é, x não é uma variável aleatória. Para problemas que consideram ambas aleatórias, pode-se adotar o método de Gauss-Helmert (conhecida na comunidade geodésica como método combinado — *mixed-model*). [Rofatto et al. \(2019\)](#) faz considerações importantes sobre as incertezas do estimador dos mínimos quadrados quando y e x são tratadas como variáveis aleatórias.

As incertezas (e_i) foram geradas artificialmente por meio de um gerador de números (pseudo)aleatórios, tópico de nossa próxima seção. Foi atribuída a distribuição normal multivariada para representar as incertezas de y_i , i.e. $e_i \sim \mathcal{N}(0, 1)$. Foram adicionados 2 outliers propositalmente nas observáveis y_1 e y_4 . As fases mostradas na Figura (1) são explicadas a seguir:

1. Primeiro, os parâmetros do modelo são determinados no processo de estimação por mínimos quadrados. O modelo inicial encontrado é mostrado em linhas preta na Figura (1). As estatísticas de teste para cada observação (mostradas em azul) são calculadas com base no vetor dos resíduos dos mínimos quadrados, denotado por w_i .
2. Posteriormente, o valor extremo da estatística de teste é localizado (sinalizado em vermelho), denotado por $\max-w$;
3. Testa-se sua significância de acordo com o valor crítico (limiar). Se a máxima estatística exceder o valor crítico, rejeita-se a hipótese nula, ou seja, identificado o outlier de contaminar o conjunto de observações, retira-se o mesmo do modelo (marcado com “**X**”).
4. Repete-se o processo até que todas os outliers suspeitos sejam identificados e devidamente removidos, em um processo iterativo de estimação, detecção, identificação e remoção do outlier (adaptação do modelo).

Nota-se por meio da Figura (1) que a redundância global do modelo ($r = n - u$) reduz a $n - 1$ a medida que um outlier é eliminado. Como consequência, em cada iteração um novo modelo sob a hipótese nula é selecionado e testado contra as $n - 1$ hipóteses

Figura 1 – Exemplo de aplicação do *Data-Snooping* em um modelo de regressão linear.



Fonte: Produzido pelos autores.

alternativas restantes. Em outras palavras, a hipótese alternativa identificada, digamos $\mathcal{H}_A^{(i)}$, se torna a nova hipótese nula e $\hat{x}^{(i)}$ é fornecido como estimador de \mathbf{x} .

Nota-se que o *Data-Snooping* é, portanto, um procedimento iterativo de eliminação de outlier. Aqui, nós usamos o termo *Data-Snooping Iterativo* (DSI), dado por Teunissen (2006, p. 135), para diferenciar do *Data-Snooping*(DS) originalmente proposto por Baarda (1968). O DSI é frequentemente designado por *Data-Snooping* (MONICO, 2008; GHILANI, 2017; LEICK; RAPOPORT; TATARNIKOV, 2015). Veremos na próxima seção que o DS original não é iterativo, e o teste envolvido com ele fixa a observação suspeita de ser um outlier (LEHMANN, 2012; YANG et al., 2013).

Como mostrado na Figura (1), os dois outliers intencionalmente gerados foram corretamente identificados e eliminados do modelo. Em uma situação real, entretanto, pode-se cometer erros na tomada de decisão, pois o DSI é uma estratégia que se utiliza de teste estatístico, ou seja, que possui níveis de probabilidade associados.

As decisões do DSI dependem do vetor dos resíduos estimados e de sua respectiva matriz de covariância. Na presença de um outlier, ou seja quando $\mathcal{H}_A^{(i)}$ é verdadeira, os

níveis de probabilidade do DSI são decompostas nas seguintes classes:

- \mathcal{P}_{CI} (*correct identification*): probabilidade de identificar e remover corretamente o outlier.
- \mathcal{P}_{MD} (*missed detection*): probabilidade de não detectar, ou seja, probabilidade de cometer falso negativo (Erro Tipo II).
- \mathcal{P}_{WE} (*wrong exclusion*): probabilidade de identificar e remover uma observação, enquanto o outlier permanece nos dados, ou seja, é a probabilidade de cometer o Erro Tipo III.
- \mathcal{P}_{over+} (*over-identification positive*): probabilidade de identificar e remover corretamente o outlier e outras observações;
- \mathcal{P}_{over-} (*over-identification negative*): probabilidade de identificar e remover mais de uma observação como sendo outlier, enquanto o outlier permanece nos dados observados.
- \mathcal{P}_{ol} (*statistical overlap*): probabilidade de detectar simultaneamente duas (ou mais) observações como outliers. Isso ocorre nos casos em que uma hipótese alternativa tem a mesma distribuição que outra. Em outras palavras, essas hipóteses não podem ser distinguidas; ou seja, elas são inseparáveis e não é possível identificar um outlier, embora a hipótese nula tenha sido corretamente rejeitada.

Portanto, a correta identificação \mathcal{P}_{CI} é o complemento da soma de todos os outros níveis de probabilidade, ou seja:

$$\mathcal{P}_{CI} = 1 - (\mathcal{P}_{MD} + \mathcal{P}_{WE} + \mathcal{P}_{over+} + \mathcal{P}_{over-} + \mathcal{P}_{ol}) \quad (1.2)$$

1.1 Justificativa

A modelagem da confiabilidade de um sistema, que opera sob o DSI, consiste em estimar o menor outlier identificável de cada observação (MIB – Minimal Identifiable Bias), com certo nível de probabilidade, e descrever sua influência sobre os parâmetros estimados do modelo, quando não identificado. Em termos de identificação, portanto, a confiabilidade interna descreve a capacidade do modelo em identificar o outlier, e a confiabilidade externa descreve a influência do MIB sobre os parâmetros do modelo, quando não identificado. A desconsideração da avaliação da confiabilidade pode conduzir a um modelo cujos parâmetros são enganosos ou equivocados. A seguir fornecemos três situações que exemplificam as consequências de não considerar a análise de confiabilidade:

- Devido à falta do planejamento de uma rede geodésica para fins de monitoramento de estrutura, pode-se concluir que uma barragem não está sendo rompida, enquanto de fato está.
- No controle de qualidade de redes geodésicas implantadas para fins de cadastro territorial, pode-se identificar e excluir erroneamente uma observação. Como resultado, a geometria da rede ficará empobrecida e comprometerá a confiabilidade dos trabalhos que dependam dela.
- Na aviação, em particular durante as fases críticas de aproximação e aterrissagem de voo, é crucial monitorar a integridade, que diz respeito à confiança que um usuário ou operador pode ter no serviço de navegação (TEUNISSEN; IMPARATO; TIBERIUS, 2017). Tal confiança é testada pelo risco à integridade. Este é a probabilidade do erro posicional exceder uma certa tolerância sem ser detectado. O operador deveria ter sido advertido sobre a falha, porém não foi. As informações de posição fornecidas são equivocadas e enganosas, pois não estão corretas se forem considerados os limites especificados e, como o usuário ou operador não são advertidos, não ficarão cientes da situação potencialmente perigosa. A aeronave está em outro lugar, diferente daquele que o sistema de navegação indica estar.

Por meio dos três exemplos acima, é possível notar que a confiabilidade está diretamente e estritamente associada à segurança. Uma falha não identificada no sistema pode colocar em risco a vida das pessoas e gerar prejuízos econômicos e socioambientais. Por exemplo, o rompimento da barragem de Fundão, localizada no município de Mariana/MG, que foi considerada um desastre ambiental sem precedentes na história do Brasil. E, mais recentemente, a barragem em Brumadinho/MG, que entrou em colapso e rompeu-se, deixando um rastro de destruição, contaminação e mortes (ver, Figura 2). Talvez desastres como estes poderiam ter sido evitados com a implantação de uma rede geodésica confiável.

Observa-se pela Equação (1.2) que a modelagem da confiabilidade requer a descrição das probabilidades envolvidas com o DSI. Entretanto, estimar os níveis de probabilidade associado com DSI é praticamente impossível por aqueles métodos analíticos usualmente empregados em procedimentos mais simples, por exemplo o DS original, cuja formulação é baseado em uma única hipótese alternativa. Por essa razão, uma rigorosa e completa teoria da confiabilidade para o DSI não estava disponível até o momento.

Embora grandes avanços tenham ocorrido em meados da década de 1970, como os computadores baseados em microprocessadores, Baarda tinha uma desvantagem: a tecnologia de sua época era insuficiente para que se utilizassem técnicas computacionais inteligentes. Hoje o cenário computacional é completamente diferente da época da teoria da confiabilidade de Baarda. Aqui, seguindo a tendência atual da ciência moderna, usamos a computação inteligente e estendemos a teoria da confiabilidade para o DSI.

Figura 2 – Resultado do rompimento das barragens em Brumadinho/MG (a) e Mariana/MG (b).



(a) Resultado do rompimento da barragem em Brumadinho/MG



(b) Resultado do rompimento da barragem em Mariana/MG

Fonte (a): <https://en.wikipedia.org/wiki>. Fonte (b): <http://prp.wpcdn.cloud>.

1.2 Objetivos Geral e Específicos

O objetivo geral da tese é estender a teoria da confiabilidade proposta originalmente por Baarda (1968). Para comprovar o objetivo geral, cinco objetivos específicos foram formulados, como segue:

- a) Avaliar a acurácia do Método Monte Carlo para a modelagem da confiabilidade.
- b) Investigar sob quais circunstâncias a dependência entre os resíduos estimados afeta os valores críticos associados à estatística de teste do DSI.
- c) Testar se o aumento do nível de significância conjunto, ou redução do valor crítico, melhora a identificabilidade do outlier.
- d) Investigar os mecanismos que associam os níveis de probabilidade do DSI com as medidas analíticas de controlabilidade das observações.
- e) Avaliar o efeito do tipo e quantidade de injunções sobre os indicadores de sensibilidade do DSI.

1.3 Estrutura da Tese

Esta Tese de Doutorado, intitulada **Uma Extensão à Teoria da Confiabilidade em Geodésia**, foi desenvolvida entre Março de 2017 e Março de 2020 no Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia (CEPSRM), da Universidade Fede-

ral do Rio Grande do Sul (UFRGS), em colaboração com o Vizlab (Advanced Visualization and GeoInformatics Lab), da Universidade do Vale do Rio dos Sinos (Unisinos).

O trabalho está dividido em cinco capítulos. Eles são descritos a seguir:

- **1: Introdução.** O objetivo do primeiro capítulo é contextualizar o problema de pesquisa. Ele foi dividido em duas seções, como segue:
 - **1.1: Justificativa.** Este tópico fornece ao leitor as razões que justificam a proposta do estudo.
 - **1.2: Objetivos Geral e Específicos.** O objetivo geral da tese é estender a teoria da confiabilidade proposta originalmente por [Baarda \(1968\)](#). Para atingir o objetivo geral, primeiro analisamos a qualidade do Método de Monte Carlo em termos de acurácia, e formulamos quatro hipóteses, cada qual com seu objetivo específico.
- **2: Fundamentação Teórica.** No segundo capítulo, apresentamos uma revisão da literatura sobre os principais elementos que sustentam esta pesquisa. O objetivo deste capítulo é mostrar a evolução cronológica da teoria da confiabilidade, desde o trabalho pioneiro de Baarda até o presente momento. Por essa razão, este capítulo foi dividido em quatro seções, como segue:
 - **2.1: Teoria da Confiabilidade Convencional.** Neste tópico, apresentamos a origem da teoria da confiabilidade.
 - **2.2: Teoria da Confiabilidade no contexto de Múltiplas Hipóteses Alternativas.** Aqui, a teoria da confiabilidade convencional é estendida para o caso de termos múltiplas hipóteses alternativas. O problema do teste de múltiplas hipóteses ocorre quando vários testes de hipóteses individuais são considerados simultaneamente. Esse é o caso do Data-Snooping. Nesta seção, apresentamos esse importante problema e mostramos que a teoria possui certas limitações que podem ser superadas quando temos a nossa disposição computadores superpotentes.
 - **2.3: Uma extensão à Teoria da Confiabilidade.** Nesta seção, demonstramos que o DSI não é somente um procedimento para identificação de outliers, mas uma combinação de estimação e teste. Demonstramos que as estimativas dos parâmetros incógnitos dependem dos níveis de probabilidade do DSI.

- **2.4: Método de Monte Carlo no contexto da Geodésia.** Como os níveis de probabilidade do DSI são analiticamente intratáveis, recorreremos ao método de Monte Carlo (MMC). Nesta seção, apresentamos um breve histórico do MMC, bem como suas aplicações e possibilidades.
- **3: Material e Métodos.** Para alcançar o objetivo geral, dois procedimentos computacionais distintos, porém complementares foram propostos, como segue:
 - **3.1: Procedimento computacional para obtenção do valor crítico da estatística $max-w$.** A estatística de teste do DSI possui sua própria distribuição de probabilidade. Conseqüentemente, os valores críticos não estão disponíveis para o DSI. Com a finalidade de superar este problema, recorreremos ao Método de Monte Carlo para construir a distribuição de probabilidade do teste envolvido no DSI.
 - **3.2: Procedimento computacional para estimação dos níveis de probabilidade do Data-Snooping Iterativo.** Nesta seção, apresentamos um procedimento computacional para estimar as probabilidades de sucesso e de falha do DSI, na presença de um outlier nos dados observados. Os valores críticos adotados nesta etapa são aqueles obtidos de acordo com a metodologia descrita na seção anterior deste capítulo.
- **4: Experimentos e Resultados.** Neste capítulo, nós apresentamos os resultados na forma de artigos. Eles foram divididos em cinco tópicos de acordo com os objetivos específicos levantados no Capítulo 1, Seção 1.2. Os tópicos são descritos a seguir:
 - **4.1: Avaliar a acurácia do Método de Monte Carlo.** Até o presente momento, a escolha do número de experimentos Monte Carlo tem sido avaliada somente em função da precisão. Assim, levanta-se uma questão: como podemos encontrar um número ótimo de experimentos Monte Carlo em termos de acurácia?
 - **4.2: Avaliação do valor crítico obtido por meio do Método de Monte Carlo.** O objetivo foi verificar sob quais circunstâncias a dependência entre os resíduos estimados afeta os valores críticos associados à estatística de teste do Data-Snooping.
 - **4.3: Análise do efeito do valor crítico controlado sobre os níveis de probabilidade do Data-Snooping.** Aqui, o objetivo foi testar se o aumento

do nível de significância conjunto, ou redução do valor crítico, melhora a identificabilidade do outlier.

- **4.4: Estudo da relação entre as medidas analíticas de controlabilidade e os níveis de probabilidade do Data-Snooping.** O objetivo foi estudar a associação entre as medidas analíticas de controlabilidade das observações e os níveis de probabilidade do Data-Snooping, na presença de um outlier dos dados observados.
- **4.5: Avaliação dos efeitos das injunções absolutas e relativas sobre a sensibilidade do Data-Snooping.** Por último, o objetivo foi avaliar o efeito do tipo e quantidade de injunção sobre os indicadores de sensibilidade do Data-Snooping.
- **5: Considerações Finais.** Neste capítulo, trazemos alguns aspectos sobre a extensão da teoria da confiabilidade. Esse capítulo foi dividido em dois tópicos, como segue:
 - **5.1: Principais Conclusões.** Com base em nossas evidências (métodos e resultados), acrescidas de informações da literatura, trazemos as principais conclusões neste tópico.
 - **5.2: Perspectivas Futuras.** Por fim, encerramos a Tese, trazendo possibilidades de desdobramentos para pesquisas futuras.

Nós também fornecemos dois Apêndices com o objetivo de auxiliar aqueles leitores que não estão familiarizados com o tema da Tese. No Apêndice (A), revisamos brevemente os principais conceitos envolvidos na teoria da estimação, e no Apêndice (B), apresentamos os principais elementos envolvidos nos testes de hipóteses em modelos lineares (ou linearizados).

2 Fundamentação Teórica

O objetivo deste capítulo é mostrar a evolução cronológica da teoria da confiabilidade, desde o trabalho pioneiro de [Baarda \(1968\)](#) até o presente momento. O leitor irá observar que, como em qualquer área do conhecimento, os avanços na teoria da confiabilidade é decorrente da evolução cronológica dos computadores. Embora as estatísticas de teste para detecção e identificação de outliers existam há mais de oito décadas, a descrição da confiabilidade para o caso do DSI somente foi possível graças à disponibilidade de microcomputadores de baixo-custo, com poder computacional suficiente para gerar artificialmente variáveis aleatórias de interesse. Nós demonstramos que o DSI não é somente teste, mas uma combinação de estimação e teste e, portanto, os resultados finais (parâmetros incógnitos do modelo) também irão depender dos níveis de probabilidade associados com o DSI. Em outras palavras, o DSI pode ser tratado como um estimador, e não somente como um procedimento de teste.

2.1 Teoria da Confiabilidade Convencional

Aqui restringimos para o caso de observáveis com distribuição normal, embora algumas pesquisas já sinalizam a não normalidade dos dados em algumas aplicações geodésicas ([LEHMANN, 2015](#); [ROFATTO et al., 2019](#)). Nesta Tese, o modelo estocástico (modelo de incerteza) das observáveis não está sujeito a discussão ou decisão. Parti-se do pressuposto que o modelo estocástico é bem conhecido e definido. Neste caso, assumimos que o modelo de incerteza sob \mathcal{H}_0 seja dado pela distribuição normal multivariada, ou seja:

$$\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_e). \quad (2.1)$$

em que $\mathbf{e} \in \mathbb{R}^{n \times 1}$ é o vetor de incertezas com seus componentes descritos na matriz covariância $\mathbf{Q}_e \in \mathbb{R}^{n \times n}$.

A hipótese nula \mathcal{H}_0 , que também é chamada de hipótese de teste, corresponde a um modelo matemático supostamente válido que descreve os dados observados. Quando assumido como “verdadeiro”, métodos de estimação, normalmente na abordagem dos mínimos quadrados, são aplicados para estimar os parâmetros do modelo. Por outro lado, pode ser proposto um modelo alternativo, denotado por \mathcal{H}_A , quando houver dúvidas sobre o nível de confiabilidade do modelo em \mathcal{H}_0 .

As duas hipóteses compostas sobre a esperança de \mathbf{y} são dadas como:

$$\mathcal{H}_0 : \mathbb{E}\{\mathbf{y}\} = \mathbf{A}\mathbf{x} + \mathbb{E}\{\mathbf{e}\} = \mathbf{A}\mathbf{x}; \mathbb{D}\{\mathbf{y}\} = \mathbf{Q}_e \quad (2.2)$$

versus

$$\mathcal{H}_A : \mathbb{E}\{\mathbf{y}\} = \mathbf{A}\mathbf{x} + \mathbf{C}\nabla + \mathbb{E}\{\mathbf{e}\} = \begin{pmatrix} \mathbf{A} & \mathbf{C} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \nabla \end{pmatrix} \quad (2.3)$$

em que $\mathbb{E}\{\cdot\}$ é o operador de esperança matemática, $\mathbb{D}\{\cdot\}$ é o operador de variância, $\mathbf{y} \in \mathbb{R}^{n \times 1}$ é o vetor das observáveis, $\mathbf{A} \in \mathbb{R}^{n \times u}$ é a matriz *design* de posto completo u , $\mathbf{x} \in \mathbb{R}^{u \times 1}$ é o vetor dos parâmetros desconhecidos, $\mathbf{e} \in \mathbb{R}^{n \times 1}$ é o vetor das incertezas e $\mathbf{Q}_e \in \mathbb{R}^{n \times n}$ é matriz covariância (simétrica positiva-definida) das observáveis \mathbf{y} .

Na hipótese alternativa em (2.3), q parâmetros adicionais (desconhecidos) no vetor $\nabla \in \mathbb{R}^{q \times 1}$ são relacionados com a esperança de \mathbf{y} por meio da matriz $\mathbf{C} \in \mathbb{R}^{n \times q}$. A matriz \mathbf{C} é assumida ser de posto completo, ou característica completa, i.e. $\text{car}(\mathbf{C}) = q$. As colunas das matrizes \mathbf{A} e \mathbf{C} também são assumidas linearmente independentes, i.e. $\text{car}(\mathbf{A} \ \mathbf{C}) = u + q$. A matriz \mathbf{C} prescreve como os efeitos não modelados se traduzem nas observações individuais, ou seja, em todos os elementos do vetor \mathbf{y} .

O número de parâmetros no vetor ∇ da hipótese alternativa em (2.3) pode variar de 1 a $n - u$, i.e. $1 \leq q \leq n - u$. Quando o número de parâmetros adicionais é igual à redundância do sistema, i.e. $q = n - u$, um teste global do modelo (*overall model test*) é realizado. Neste caso extremo, a redundância da hipótese alternativa é $r = n - (u + q) = n - u - n + u = 0$. Logo, não há necessidade de especificar a hipótese alternativa. Outro caso especial é quando tomamos $q = 1$, ou seja, estendemos o modelo da hipótese nula com um parâmetro adicional. Neste caso, tem-se o teste com natureza local (*local model test*).

No procedimento *Data-Snooping* (DS) proposto originalmente por Baarda (1968), o teste local do modelo (*local model test*) é aplicado com o objetivo de localizar um possível outlier nos dados observados. Para rastrear as observações, a fim de identificar o outlier, são formuladas n hipóteses alternativas no DS, como segue:

$$\mathcal{H}_A^{(i)} : \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{c}_i \nabla_i + \mathbf{e} = \begin{pmatrix} \mathbf{A} & \mathbf{c}_i \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \nabla_i \end{pmatrix} + \mathbf{e}, \forall i = 1, \dots, n \quad (2.4)$$

com $\mathbf{c}_i \in \mathbb{R}^{n \times 1}$ um vetor canônico, cujos elementos são “0” e “1”, em que “1” fornece a posição de um possível outlier nos dados observados, enquanto “0” é atribuído para as demais observações; e ∇_i é um escalar que corresponde à magnitude do outlier. Restringimos para o caso em que $\begin{pmatrix} \mathbf{A} & \mathbf{c}_i \end{pmatrix} \in \mathbb{R}^{n \times (u+1)}$ também possuem posto completo $u + 1$, ou seja, todas as colunas são linearmente independentes. A redundância no caso de uma única hipótese alternativa $\mathcal{H}_A^{(i)}$ é $n - (u + q) = n - u - 1 = n - \text{car}(\mathbf{A} \ \mathbf{c}_i)$.

Nota-se pela Equação (2.4) que as hipóteses alternativas são testadas individualmente contra o modelo padrão em \mathcal{H}_0 . Por exemplo, se deseja testar uma observação

na i -ésima posição de \mathbf{y} , tem-se $\mathbf{c}_i = (0 \ 0 \ 0 \ \dots \ 1^{i\text{-ésima}} \ 0 \ \dots \ 0)^T$. Como se tem n hipóteses alternativas, testa-se cada uma individualmente, variando a posição i do elemento “1” de 1 a n no vetor \mathbf{c}_i .

A estatística de teste associado com o DS é conhecida como w -test, e ela é dada por:

$$\mathbf{w}_i = \frac{\mathbf{c}_i^T \mathbf{Q}_e^{-1} \hat{\mathbf{e}}}{\sqrt{\mathbf{c}_i^T \mathbf{Q}_e^{-1} \mathbf{Q}_{\hat{\mathbf{e}}} \mathbf{Q}_e^{-1} \mathbf{c}_i}}, \forall i = 1, \dots, n \quad (2.5)$$

com

$$\begin{cases} \hat{\mathbf{e}} = \mathbf{P}_A^\perp \mathbf{y} \\ \mathbf{Q}_{\hat{\mathbf{e}}} = \mathbf{P}_A^\perp \mathbf{Q}_e \mathbf{P}_A^{\perp T} \end{cases} \quad (2.6)$$

Na Equação (2.5), tem-se os seguintes elementos: $\hat{\mathbf{e}} \in \mathbb{R}^{n \times 1}$ o vetor dos resíduos estimados; $\mathbf{Q}_{\hat{\mathbf{e}}} \in \mathbb{R}^{n \times n}$ a matriz covariância dos resíduos estimados; $\mathbf{P}_A = \mathbf{A}(\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W}$ e $\mathbf{P}_A^\perp = \mathbf{I} - \mathbf{P}_A$ sendo dois projetores ortogonais; \mathbf{I} é uma matriz identidade, i.e. $\mathbf{I} \in \mathbb{R}^{n \times n}$. O projetor \mathbf{P}_A projeta sobre o espaço coluna da matriz \mathbf{A} (denotado por $\mathcal{R}(\mathbf{A})$) e na direção de seu complemento ortogonal (i.e. $\mathcal{R}(\mathbf{A})^\perp$), enquanto que \mathbf{P}_A^\perp projeta sobre $\mathcal{R}(\mathbf{A})^\perp$ e na direção de $\mathcal{R}(\mathbf{A})$. Para algumas propriedades úteis desses dois projetores, consulte Teunissen (2006). $\mathcal{R}(\cdot)$ denota o espaço coluna de uma matriz. Importante destacar que projetor \mathbf{P}_A^\perp é conhecido como matriz redundância, denotado aqui por \mathbf{R} , i.e. $\mathbf{R} = \mathbf{I} - \mathbf{A}(\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W}$. A diagonal da matriz \mathbf{R} é composto por elementos conhecidos como número de redundância local, denotado por r_i . Para esta formulação, a matriz covariância \mathbf{Q}_e das observáveis é conhecida e, portanto, pode-se usar o (BLUE) tomando a matriz dos pesos \mathbf{W} como a inversa da \mathbf{Q}_e , ou seja, $\mathbf{W} = \mathbf{Q}_e^{-1}$.

A estatística de teste em (2.5) segue distribuição normal padrão. Neste caso, tem-se:

$$\mathcal{H}_0 : \mathbf{w}_i \sim \mathcal{N}(0, 1); \quad \mathcal{H}_A^{(i)} : \mathbf{w}_i \sim \mathcal{N}(\nabla w_i, 1) \quad (2.7)$$

com $\nabla w_i = (\mathbf{c}_i^T \mathbf{Q}_e^{-1} \mathbf{Q}_{\hat{\mathbf{e}}} \mathbf{Q}_e^{-1} \mathbf{c}_i)^{1/2} \nabla_i$. O parâmetro ∇w_i está relacionado com o parâmetro de não-centralidade da seguinte forma: $\lambda_{0(i)} = \nabla w_i^2$, em que:

$$\lambda_{0(i)} = \nabla_i^T \mathbf{c}_i^T \mathbf{P}_A^{\perp T} \mathbf{Q}_e^{-1} \mathbf{P}_A^\perp \mathbf{c}_i \nabla_i, \forall i = 1, \dots, n \quad (2.8)$$

No caso do procedimento *Data-Snooping*, a região de aceitação do w -test é dada por:

$$\mathcal{A}_{\alpha_0} = \{\mathbf{y} \in \mathbb{R}^n \mid |\mathbf{w}_i| \leq \sqrt{k_0}\} \quad (2.9)$$

com k_0 sendo o valor crítico tabelado na distribuição qui-quadrado, para $q = 1$ grau de liberdade à um nível de significância de α_0 . Por exemplo, para um $\alpha_0 = 0.01$, obtém-se $\sqrt{k_0} = 2.576$. Neste caso, se $|\mathbf{w}_i| > 2.576$ para uma dada observação y_i , pode-se rejeitar a hipótese nula, ou seja, $y_i \notin \mathcal{A}_{\alpha_0}$ ¹.

¹ NOTA: O subscrito $(_0)$ indica que as quantidades associadas são para o caso em que há apenas uma única hipótese alternativa envolvida no teste.

Ao aplicar o DS em sua forma original, dois tipos de decisões falsas podem ocorrer:

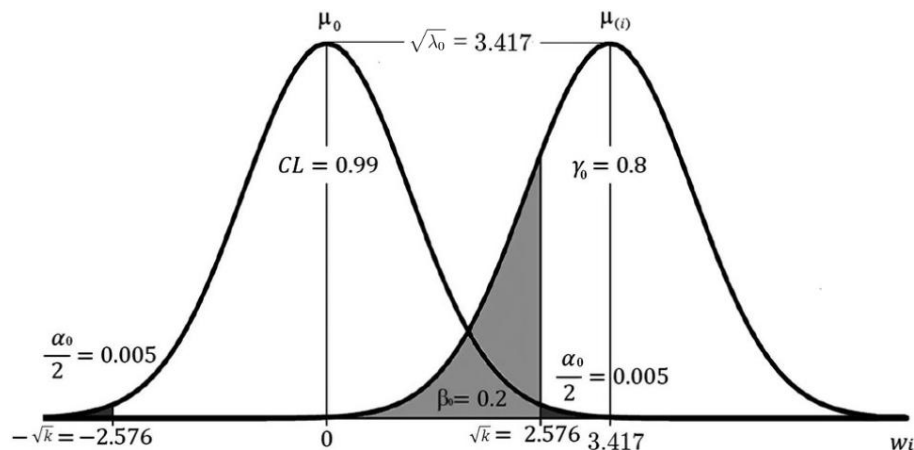
- Erro tipo I — α_0 : probabilidade de rejeitar a hipótese nula \mathcal{H}_0 , quando na realidade é verdadeira. No contexto de outlier, é a probabilidade de detectar um outlier, quando na verdade, não existe outlier.
- Erro tipo II — β_0 : é a probabilidade de rejeitar a hipótese alternativa $\mathcal{H}_A^{(i)}$, quando na realidade ela é verdadeira, ou seja, é a probabilidade de não detectar o outlier, quando de fato a observação testada é um outlier.

Os complementares de α_0 e β_0 são, respectivamente, o intervalo de confiança $I.C. = 1 - \alpha_0$ e poder do teste $\gamma_0 = 1 - \beta_0$.

A presença de um outlier causa perda da propriedade de imparcialidade do estimador de mínimos quadrados, pois a esperança matemática do erro do estimador será maior que zero. Além disso, o estimador deixará de ser de variância mínima. Em termos de distribuição de probabilidade, o outlier desloca a esperança matemática da f.d.p. sob a hipótese nula \mathcal{H}_0 para a esperança matemática da f.d.p. sob a hipótese alternativa $\mathcal{H}_A^{(i)}$. Esse efeito é conhecido como deslocamento médio do modelo padrão, ou inglês *mean-shift-model*. Observe que a esperança matemática da $\mathcal{H}_A^{(i)}$ é exatamente a raiz quadrada do parâmetro de não-centralidade, ou seja, ∇w_i .

Como pode ser visto no exemplo da Figura (3), a raiz quadrada do parâmetro de não-centralidade, denotado por $\sqrt{\lambda_0}$, depende de α_0 , β_0 e dos graus de liberdade do teste q . Como já visto, no caso do DS, temos um teste local, ou seja, $q = 1$. Com efeito, torna-se possível quantificar a menor magnitude de um outlier que leva à rejeição do modelo sob a hipótese nula. Seguindo essa linha de raciocínio, nasce a teoria de confiabilidade proposta por Baarda (1968).

Figura 3 – Exemplo da relação entre o parâmetro de não-centralidade e os níveis de probabilidade associados com o Data-Snooping.



Fonte: Produzido pelos autores.

Com base nos erros na tomada de decisão do DS, [Baarda \(1968\)](#) propôs o conceito de menor outlier detectável, traduzido do nome original Minimal Detectable Bias (MDB), termo dado por [Teunissen \(1990\)](#). O MDB de uma dada hipótese alternativa $\mathcal{H}_A^{(i)}$ é a menor magnitude de um outlier que pode levar a rejeição da hipótese nula a um certo nível de probabilidade. O MDB descreve a confiabilidade interna do modelo, enquanto que a propagação dos MDBs para os parâmetros incógnitos do modelo descreve a confiabilidade externa ([VANÍČEK; CRAYMER; KRAKIWSKY, 2001](#)). Em outras palavras, a confiabilidade interna descreve a capacidade do modelo em detecção outlier, e a confiabilidade externa descreve a influência do MDB sobre os parâmetros incógnitos do modelo, quando não detectado.

Sob a condição da hipótese nula \mathcal{H}_0 (sem presença de outlier), a estatística de teste em (2.5) segue distribuição normal padrão. Por outro lado, a existência de um único outlier na i -ésima posição do conjunto de dados faz com que a esperança de matemática de w_i se torne igual à raiz quadrada do parâmetro de não-centralidade da distribuição χ^2 para $q = 1$ (veja as equações 2.7 e 2.8). Neste caso, o parâmetro de não-centralidade $\lambda_{0(i)}$ é dado como:

$$\mathbb{E}\{w_i\} = \sqrt{\lambda_{0(i)}} = \sqrt{\mathbf{c}_i^T \mathbf{Q}_e^{-1} \mathbf{Q}_{\hat{\mathbf{e}}_0} \mathbf{Q}_e^{-1} \mathbf{c}_i \nabla_i^2} \quad (2.10)$$

O parâmetro de não-centralidade pode ser calculado com base nos níveis de probabilidades previamente estipulados α_0 e β_0 (ou $I.C.$ e γ_0 , respectivamente). Uma forma de obter esses valores computacionalmente é fornecido por [Aydin e Demirel \(2004\)](#).

O termo dentro da raiz quadrada da Equação (2.10), i.e. $\mathbf{c}_i^T \mathbf{Q}_e^{-1} \mathbf{Q}_{\hat{\mathbf{e}}_0} \mathbf{Q}_e^{-1} \mathbf{c}_i$, é um escalar. Ele tem uma relação direta com a incerteza, pois fornece a desvio-padrão do outlier, ou seja, $\sigma_{\hat{\nabla}_i} = \sqrt{(\mathbf{c}_i^T \mathbf{Q}_e^{-1} \mathbf{Q}_{\hat{\mathbf{e}}_0} \mathbf{Q}_e^{-1} \mathbf{c}_i)^{-1}}$. Logo, pode-se escrever:

$$|\nabla_i| = MDB_{0(i)} = \sigma_{\hat{\nabla}_i} \sqrt{\lambda_{0(i)}}, \quad \forall i = 1, \dots, n \quad (2.11)$$

em que $|\nabla_i| = MDB_{0(i)}$ expressa o menor outlier que pode ser detectado pelo procedimento DS unidimensional (ou seja, uma única hipótese alternativa, $q = 1$), segundo os níveis de probabilidades α_0 e β_0 que forem estipulados. Baarda referiu-se aos menores outliers detectáveis $|\nabla_i|$ como “boundary values” (em holandês: grenswaarden).

A influência de um MDB não detectado sob os parâmetros do modelo pode ser calculado para cada MDB individual. Neste caso, tem-se os seguintes resultados para confiabilidade externa:

$$\begin{cases} \hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{Q}_e^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Q}_e^{-1} \mathbf{c}_i |\nabla_i| \\ \hat{\mathbf{y}} = P_A \mathbf{c}_i |\nabla_i| \\ \hat{\mathbf{e}} = P_A^\perp \mathbf{c}_i |\nabla_i| \end{cases} \quad (2.12)$$

em que $\hat{\mathbf{x}} \in \mathbb{R}^{u \times 1}$ produz o vetor das estimativas dos parâmetros incógnitos do modelo, sob o efeito de um outlier não detectado; $\hat{\mathbf{y}} \in \mathbb{R}^{n \times 1}$ produz o vetor das observações ajustadas

sob o efeito de um outlier não detectado; e $\hat{\boldsymbol{\epsilon}} \in \mathbb{R}^{n \times 1}$ são os resíduos estimados sob o efeito de um outlier não detectado.

Importante notar que cada MDB irá produzir estimativas diferentes. O pior caso é aquele que fornece a maior influência sob os parâmetros do modelo. Para detalhes veja [Matsuoka et al. \(2020\)](#).

Observe que a rejeição da hipótese nula, automaticamente implica a aceitação da hipótese alternativa. Portanto, originalmente, o *Data snooping* é para fins de detecção de outlier, e não identificação.

O problema original foi proposto considerando a observação fixa. Em outras palavras, a hipótese alternativa foi formulada de acordo com a seguinte afirmação: **uma observação específica é um outlier**. Na prática, entretanto, um outlier é sempre desconhecido.

Do ponto de vista prático, não sabemos qual observação é um outlier. Portanto, uma hipótese alternativa mais apropriada seria: **existe pelo menos um outlier no vetor das observações**. Ao fazer tal afirmação, recaímos sobre um problema de múltiplas hipóteses alternativas.

2.2 Teoria da Confiabilidade no contexto de Múltiplas Hipóteses Alternativas

Agora, estamos interessados em saber qual das hipóteses alternativas pode levar à rejeição da hipótese nula com uma certa probabilidade. Isto significa testar \mathcal{H}_0 contra $\mathcal{H}_A^{(1)}, \mathcal{H}_A^{(2)}, \mathcal{H}_A^{(3)}, \dots, \mathcal{H}_A^{(n)}$, simultaneamente ([LEHMANN, 2012](#)). Esse tipo de teste é conhecido na literatura estatística como testes simultâneos de várias hipóteses nulas, ou simplesmente testes múltiplos ([HSU, 1996](#); [LEHMANN](#); [ROMANO, 2005](#); [MILLER, 1981](#); [SARKAR](#); [CHANG, 1997](#); [SIMES, 1986](#); [ROM, 2013](#); [WRIGHT, 1992](#)). Neste contexto, os métodos que lidam com múltiplas hipóteses alternativas são referenciados como Métodos de Comparações Múltiplas. Investigações sobre tais métodos são raros em aplicações geodésicas, o que, de certa forma, abre caminhos para pesquisa e aplicações.

Testes de múltiplas hipóteses é, em princípio, um conjunto de n testes de hipóteses convencionais, \mathcal{H}_0 contra $\mathcal{H}_A^{(i)}$, com as estatísticas de teste w_i calculadas por (2.5), e valores críticos k_i , para $i = 1, \dots, n$. Se em qualquer um destes n testes k_i é excedido por um w_i , então a hipótese nula \mathcal{H}_0 é rejeitada, ou seja:

$$\text{Rejeitar } \mathcal{H}_0 \text{ se } \mathbf{w}_1 > k_1 \text{ ou } \mathbf{w}_2 > k_2 \text{ ou } \mathbf{w}_3 > k_3 \text{ ou } \dots \text{ ou } \mathbf{w}_n > k_n \quad (2.13)$$

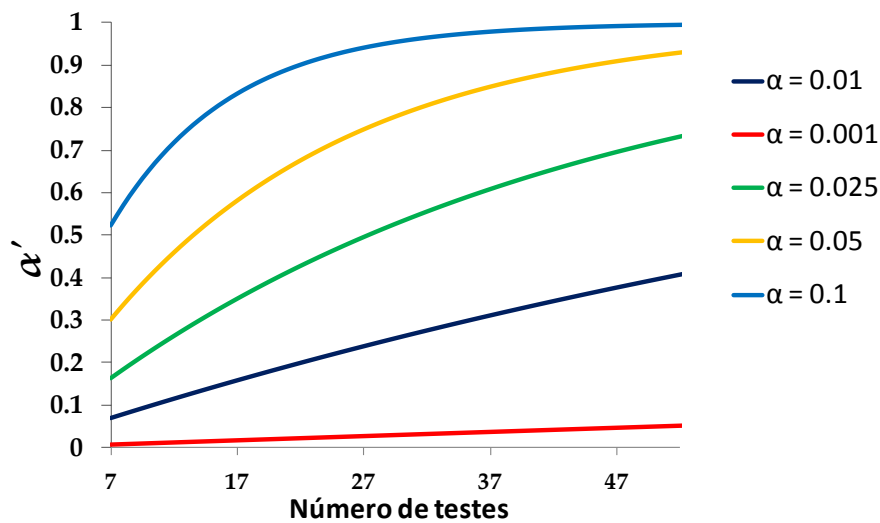
Neste caso, deve-se supor que as observações contenham outliers. Caso contrário, o modelo da hipótese nula \mathcal{H}_0 pode ser usado no processo de estimação.

O evento indesejado de rejeitar uma \mathcal{H}_0 “verdadeira” pode ocorrer em qualquer um dos n testes. Suponha que a probabilidade de rejeitar uma hipótese nula \mathcal{H}_0 erroneamente em um teste i seja α_i (conhecida como *experiment-wise error rate*), e que $\alpha_i \ll 1$. Além disso, considere que os eventos aleatórios de “rejeitar a hipótese nula verdadeira” no teste i sejam estatisticamente independentes. Então a probabilidade “total” de rejeitar a verdadeira hipótese nula \mathcal{H}_0 em múltiplos testes (conhecida como *familywise error rate* — FWER, ou taxa de erro da família dos testes) é aproximadamente (LEHMANN; LÖSLER, 2016):

$$\alpha' \approx 1 - \prod_{i=1}^n (1 - \alpha_i) = 1 - (1 - \alpha_i)^n \quad (2.14)$$

Por exemplo, considerando o nível de significância para cada teste individual $\alpha_i = 0.05$. Se executarmos 10 testes, a probabilidade de rejeitarmos pelo menos uma hipótese verdadeira é 0.4, e se fossem executados 50 testes para o mesmo $\alpha_i = 0.05$, então teríamos 0.92. A Figura (4) ilustra esse exemplo. Em estatística, o nível de significância α' na expressão (2.14) é chamado de nível de significância conjunto.

Figura 4 – Efeito do aumento do número de testes sobre (α').



Fonte: Produzido pelos autores.

Uma forma de controlar o FWER, é escolher um parâmetro de modo que o risco total de rejeitar uma verdadeira \mathcal{H}_0 seja particionado igualmente para os testes individuais. Fazendo:

$$\alpha_i := \alpha_0 \quad (2.15)$$

essa condição é satisfeita, embora não corresponda exatamente a α' .

O objetivo da expressão (2.15) é controlar adequadamente o nível de significância conjunto (α'), isto é, garantir que a “probabilidade de ao menos uma rejeição falsa” não

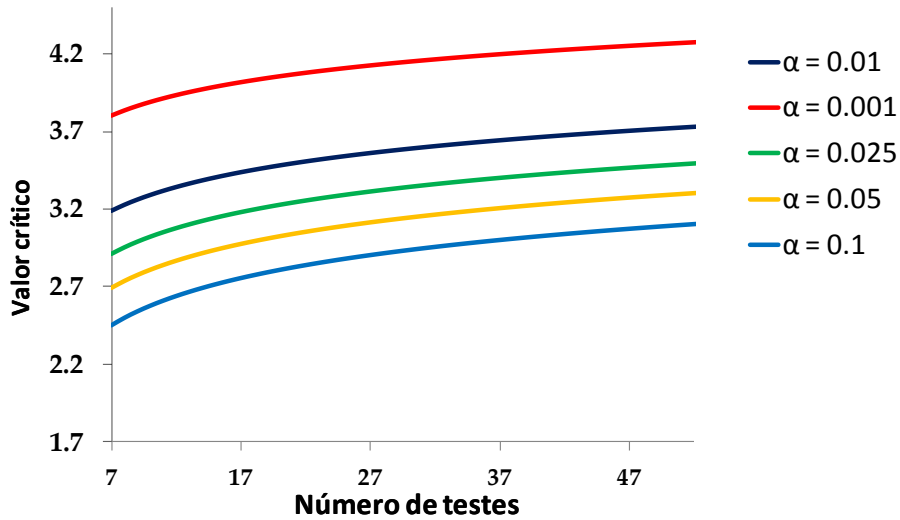
aumente com o número de testes envolvidos. A Equação (2.15) é conhecida como equação de Bonferroni (ABDI, 2006).

Os valores críticos agora são dados por:

$$k_i = F_N^{-1}(1 - \alpha_0) \approx F_N^{-1}\left(1 - \frac{\alpha'}{2n}\right) \quad (2.16)$$

em que F_N^{-1} é a inversa da função de densidade de probabilidade acumulada da distribuição normal padrão. O número “2” no numerador significa que realizamos um teste bilateral. A Figura (5) mostra os valores críticos calculado pela Equação (2.16), com o objetivo de controlar o FWER mostrado na Figura (3).

Figura 5 – Valores críticos calculados pela relação de Bonferroni.



Fonte: Produzido pelos autores.

Isso mostra que para os testes individuais, os valores críticos também serão idênticos, ou seja, $k_i = \hat{k}$. Portanto, é suficiente compará-los apenas com o máximo das estatísticas de teste. Portanto, o procedimento DS para $q = 1$ tem efeito quando retorna o maior valor absoluto entre as estatísticas de teste, denotado por $\max-w$, ou seja:

$$\max-w = \max_{i \in \{1, \dots, n\}} |w_i| \quad (2.17)$$

Agora, a regra de decisão é dada por:

$$\begin{aligned} & \text{Aceitar } \mathcal{H}_0 \text{ se } \max-w \leq \hat{k} \\ & \text{Caso contrário,} \\ & \text{Aceitar } \mathcal{H}_A^{(i)} \text{ se } \max-w > \hat{k} \end{aligned} \quad (2.18)$$

A regra de decisão em (2.18) diz que se nenhum dos n w -tests for rejeitado, então aceitamos a hipótese nula \mathcal{H}_0 . Se a hipótese nula \mathcal{H}_0 for rejeitada em qualquer um dos

n testes, será possível concluir apenas que a detecção ocorreu. Em outras palavras, se $\max-w$ for maior que um dado percentil de sua distribuição de probabilidade (i.e. um dado valor crítico \hat{k}), então existem evidências para a presença de outlier nos dados observados. Portanto, “detecção de outlier” somente nos informa se a hipótese nula \mathcal{H}_0 deve ser aceita ou não. A detecção não nos diz qual hipótese alternativa $\mathcal{H}_A^{(i)}$ teria levado a rejeição da hipótese nula \mathcal{H}_0 .

A localização da hipótese alternativa, que teria levado a rejeição da hipótese nula, é um problema de “identificação de outlier”. A identificação de outlier implica na busca para o outlier mais provável nos dados observados. Em outras palavras, busca-se descobrir dentre as estatísticas de teste w -test disponíveis aquele que produziu a estatística de máximo valor absoluto de w_i , ou seja, $\max-w$. Observe, portanto, que para ocorrer a identificação, necessariamente deve ter ocorrido a detecção.

Portanto, o *Data-snooping* é, na verdade, um importante procedimento de múltiplas hipóteses alternativas, e não um caso de uma única hipótese alternativa. Conseqüentemente, a correta rejeição da hipótese nula não implica automaticamente na correta identificação de uma hipótese alternativa. Neste caso, além do Erro Tipo I e do Erro Tipo II, temos a probabilidade de rejeitar corretamente a hipótese nula, porém identificar erroneamente a hipótese alternativa. Em outras palavras, pode-se corretamente detectar que um possível outlier se faz presente nos dados observados, porém pode-se identificar erroneamente uma observação como sendo outlier, enquanto o outlier permanece nas observações. Esse tipo de erro é conhecido como Erro Tipo III e, aqui denotado por κ_{ij} .

Geralmente em aplicações de posicionamento por satélites GNSS(Global Navigation Satellite System) e navegação, a probabilidade do Erro Tipo I é chamado de *false alarm*(\mathcal{P}_{FA}); o Erro Tipo II, de *missed detection*(\mathcal{P}_{MD}); e o Erro Tipo III, de *wrong identification*(\mathcal{P}_{WI}).

A dependência entre as hipóteses tem uma importante papel no que diz respeito à identificação. Ela é medida por meio da correlação entre as estatísticas de teste w -test. A correlação entre dois testes, por exemplo w_i e w_j , é dada por (FÖRSTNER, 1983):

$$\rho_{w_i, w_j} = \frac{\mathbf{c}_i^T \mathbf{Q}_e^{-1} \mathbf{Q}_{\hat{e}_0} \mathbf{Q}_e^{-1} \mathbf{c}_j}{\sqrt{\mathbf{c}_i^T \mathbf{Q}_e^{-1} \mathbf{Q}_{\hat{e}_0} \mathbf{Q}_e^{-1} \mathbf{c}_i} \sqrt{\mathbf{c}_j^T \mathbf{Q}_e^{-1} \mathbf{Q}_{\hat{e}_0} \mathbf{Q}_e^{-1} \mathbf{c}_j}}, \forall (i \neq j) \quad (2.19)$$

A confiabilidade no caso de múltiplas hipóteses alternativas envolve análise de separabilidade. A análise de separabilidade tem por objetivo indicar quais hipóteses alternativas podem ser distinguidas. Ela é um elemento crucial no monitoramento da integridade do posicionamento/navegação baseado em satélites GNSS(Global Navigation Satellite System), conhecido como RAIM(*Receiver Autonomous Integrity Monitoring*)(HEWITSON; WANG, 2006).

Tabela 1 – Níveis de probabilidade associado com o *Data-snooping* sob múltiplas hipóteses alternativas

Realidade <i>Desconhecida</i>	Resultado do Teste				
	\mathcal{H}_0	$\mathcal{H}_A^{(1)}$	$\mathcal{H}_A^{(2)}$	\dots	$\mathcal{H}_A^{(n)}$
\mathcal{H}_0	correta decisão $1-\alpha'$	Erro Tipo I α_1	Erro Tipo I α_2	\dots	Erro Tipo I α_n
$\mathcal{H}_A^{(1)}$	Erro Tipo II β_{10}	Correct Identificação $1-\beta_{11}$	Erro Tipo III κ_{12}	\dots	Erro Tipo III κ_{1n}
$\mathcal{H}_A^{(2)}$	Erro Tipo II β_{20}	Erro Tipo III κ_{21}	Correct Identificação $1-\beta_{22}$	\dots	Erro Tipo III κ_{2n}
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
$\mathcal{H}_A^{(n)}$	Erro Tipo II β_{n0}	Erro Tipo III κ_{n1}	Erro Tipo III κ_{n2}	\dots	Correct Identificação $1-\beta_{nn}$

Fonte: Produzido pelo autores.

Os níveis de probabilidade associado com *Data-snooping* sob múltiplas hipóteses alternativas é dada na Tabela (1).

O cenário sem presença de outliers, ou seja, quando a hipótese nula \mathcal{H}_0 é verdadeira, foi detalhado anteriormente ao descrever como controlar o FWER.

Com base no pressuposto de que um outlier esteja na i th posição dos dados observados (i.e., $\mathcal{H}_A^{(i)}$ é “*verdadeira*”), a taxa de Erro Tipo II (“*missed detection*” — \mathcal{P}_{MD}) é dada por:

$$\mathcal{P}_{MD} = \beta_{i0} = \mathcal{P} \left(\bigcap_{i=1}^n |\mathbf{w}_i| \leq \hat{k} \mid \mathcal{H}_A^{(i)} : \textit{verdadeira} \right), \quad (2.20)$$

e o tamanho do Erro Tipo III (também chamado de “*misidentification*”, denotado por \mathcal{P}_{WI}) é

$$\mathcal{P}_{WI} = \sum_{j=1}^{n-1} \kappa_{ij} = \sum_{j=1}^{n-1} \mathcal{P} \left(|\mathbf{w}_j| > |\mathbf{w}_i| \forall i, |\mathbf{w}_j| > \hat{k} (i \neq j) \mid \mathcal{H}_A^{(i)} : \textit{verdadeira} \right) \quad (2.21)$$

Por outro lado, a probabilidade de correta identificação (“*correct identification*”, denotado por \mathcal{P}_{CI}) é dado como

$$\mathcal{P}_{CI} = 1 - \beta_{ii} = \mathcal{P} \left(|\mathbf{w}_i| > |\mathbf{w}_j| \forall j, |\mathbf{w}_i| > \hat{k} (i \neq j) \mid \mathcal{H}_A^{(i)} : \textit{verdadeira} \right) \quad (2.22)$$

com

$$1 - \mathcal{P}_{CI} = \beta_{ii} = \beta_{i0} + \sum_{i=1}^n \kappa_{ij}, \textit{ for } (i \neq j) \quad (2.23)$$

Observe que a soma desses três níveis de probabilidade é exatamente 1, ou seja: $\mathcal{P}_{MD} + \mathcal{P}_{WI} + \mathcal{P}_{CI} = 1$.

A probabilidade da correta detecção \mathcal{P}_{CD} é a soma da probabilidade da correta identificação \mathcal{P}_{CI} (selecionar uma hipótese alternativa corretamente) e a probabilidade da

identificação incorreta \mathcal{P}_{WI} (selecionar uma das $n - 1$ hipóteses), i.e.,

$$\mathcal{P}_{CD} = \mathcal{P}_{CI} + \mathcal{P}_{WI} \quad (2.24)$$

A probabilidade de identificar erroneamente a hipótese alternativa é identicamente zero ($\mathcal{P}_{WI} = 0$), quando o coeficiente de correlação é exatamente zero $\rho_{w_i, w_j} = 0$. Neste caso, temos a seguinte relação:

$$\mathcal{P}_{CD} = \mathcal{P}_{CI} = 1 - \mathcal{P}_{MD} \quad (2.25)$$

A relação dada na Equação (2.25) só aconteceria se negligenciasse a natureza da dependência entre as hipóteses alternativas. Em outras palavras, esse relacionamento é válido para o caso especial de testar a hipótese nula \mathcal{H}_0 contra apenas uma única hipótese alternativa $\mathcal{H}_A^{(i)}$.

Geometricamente, a região crítica no teste de múltiplas hipóteses é maior que aquela do teste com uma única hipótese alternativa. Logo, o Erro Tipo II (ou seja, \mathcal{P}_{MD}) se torna menor para múltiplos testes (ROFATTO et al., 2020a). Isso significa que a detecção no caso de múltiplas hipóteses alternativas é mais poderosa que no caso do teste binário, ou seja:

$$\mathcal{P}_{CD} > \gamma_0 \quad (2.26)$$

Entretanto, a detecção correta \mathcal{P}_{CD} sob n testes é espalhada por todas as hipóteses alternativas, corretas ou erradas, diminuindo a probabilidade da correta identificação. Em outras palavras, a separabilidade depende da correlação entre as estatísticas de teste (ρ_{w_i, w_j}). Quanto maior é esta correlação, maior é a taxa do Erro Tipo III e menor é a probabilidade do Erro Tipo II (YANG et al., 2013). Em outras palavras, aumenta-se o poder da detecção, porém reduz a probabilidade da correta identificação de um outlier.

Observa-se também pela Equação (2.24) que a detecção não depende da identificação, o contrário não é verdade (IMPARATO; TEUNISSEN; TIBERIUS, 2019). Portanto, temos a seguinte desigualdade:

$$\mathcal{P}_{CI} \leq \mathcal{P}_{CD} \quad (2.27)$$

Embora, em teoria, seja possível descrever as funções densidade de probabilidade das decisões estatísticas associadas com o DS sob múltiplas hipóteses alternativas, na prática, o cálculo algébrico de tais funções é altamente complexo, e mesmo impossíveis de serem obtidas por um expressão analítica fechada, ou mesmo por integração numérica (PRÓSZYŃSKI, 2015). Na tentativa de superar essa limitação, Yang et al. (2013) apresentam uma teoria que estabelece os limites dos erros de decisão do DS sob múltiplas hipóteses alternativas. Com essa teoria, as probabilidades de cometer erros do tipo I, II e III podem ser controladas para que a probabilidade de identificação de um outlier possa ser

garantida ao executar o DS. Entretanto, os mesmos autores apontam que o cálculo de tais limites somente é possível por meio de técnicas de simulações computacionais. Por meio do Método de Monte Carlo, eles estendem a teoria convencional de Baarda, 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). Segundo, consideraram um único processo de estimação e teste. Neste caso, o DS é executado uma única vez. Esse não é o padrão encontrado na maioria dos softwares de processamento de dados geodésicos. Na prática, portanto, lidamos com *Data-Snooping*s em sua forma iterativa e, portanto, a probabilidade da correta identificação \mathcal{P}_{CI} depende de outras regras de decisão.

2.3 Uma Extensão à Teoria da Confiabilidade

Aqui, alguns importantes conceitos são retomados para auxiliar no desenvolvimento matemático abaixo.

Assumimos que o modelo da hipótese nula seja dado com a seguinte estrutura:

$$\mathcal{H}_0 : \mathbb{E}\{\mathbf{y}\} = \mathbf{A}\mathbf{x}; \mathbb{D}\{\mathbf{y}\} = \mathbf{Q}_e \quad (2.28)$$

com $\mathbb{E}\{\cdot\}$ sendo o operador de esperança matemática, $\mathbb{D}\{\cdot\}$ é o operador de dispersão, $\mathbf{y} \in \mathbb{R}^{n \times 1}$ é o vetor aleatório normalmente distribuído das observáveis, $\mathbf{A} \in \mathbb{R}^{n \times u}$ é a matriz *design* de posto completo u , $\mathbf{x} \in \mathbb{R}^{u \times 1}$ é o vetor dos parâmetros desconhecidos, $\mathbf{e} \in \mathbb{R}^{n \times 1}$ é o vetor das incertezas, e $\mathbf{Q}_e \in \mathbb{R}^{n \times n}$ é matriz covariância (simétrica positiva-definida) das observáveis \mathbf{y} . A redundância de \mathcal{H}_0 é $r = n - \text{car}(\mathbf{A}) = n - u$.

Sob \mathcal{H}_0 , o melhor estimador linear imparcial (BLUE) de \mathbf{x} e \mathbf{e} é dado como:

$$\hat{\mathbf{x}}_0 = \mathbf{A}^+ \mathbf{y}, \mathbf{e} \quad (2.29)$$

$$\hat{\mathbf{e}}_0 = \mathbf{R}\mathbf{y}, \quad (2.30)$$

com $\mathbf{A}^+ = (\mathbf{A}^T \mathbf{Q}_e^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Q}_e^{-1}$ sendo a inversa de \mathbf{A} nos termos do BLUE. Como já visto, $\mathbf{R} \in \mathbb{R}^{n \times n}$ é um projetor ortogonal conhecido como matriz redundância e pode ser calculada como $\mathbf{R} = \mathbf{I} - \mathbf{A}(\mathbf{A}^T \mathbf{Q}_e^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Q}_e^{-1}$, em que $\mathbf{I} \in \mathbb{R}^{n \times n}$ é a matriz identidade. A matriz redundância \mathbf{R} foi o ponto chave para o desenvolvimento do método proposto nesta Tese e será devidamente detalhada no Capítulo 3.

Como a qualidade de $\hat{\mathbf{x}}_0$ ² depende da validade da hipótese nula, é importante que se tenha suficiente confiança em \mathcal{H}_0 . Na presença de um outlier, nós assumimos que

² NOTA: O subscripto “0” é usado tanto para indicar as quantidades associadas com o teste de única hipótese alternativa como para designar as estimativas feitas sob \mathcal{H}_0 . As estimativas em $\mathcal{H}_A^{(i)}$ são denotadas pelo subscripto “i”.

seu efeito afeta a esperança (média) de \mathbf{y} . Como um outlier é sempre desconhecido, n modelos alternativos podem ser desenvolvidos. Neste caso, o modelo da hipótese nula é reparametrizado para $q = 1$, dando origem a n hipóteses alternativas, como segue:

$$\mathcal{H}_A^{(i)} : \mathbb{E}\{\mathbf{y}\} = \mathbf{A}\mathbf{x} + \mathbf{c}_i \nabla_i = \begin{pmatrix} \mathbf{A} & \mathbf{c}_i \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \nabla_i \end{pmatrix}, \forall i = 1, \dots, n \quad (2.31)$$

com $\mathbf{c}_i \in \mathbb{R}^{n \times 1}$ um vetor canônico, cujos elementos são “0” e “1”, em que “1” fornece a posição de um possível outlier nos dados observados, enquanto “0” é atribuído para as demais observações; e ∇_i é um escalar que corresponde à magnitude do outlier. Restringimos para o caso em que $\begin{pmatrix} \mathbf{A} & \mathbf{c}_i \end{pmatrix} \in \mathbb{R}^{n \times (u+1)}$ também possuem posto completo $u + 1$, ou seja, todas as colunas são linearmente independentes. A redundância no caso de uma única hipótese alternativa $\mathcal{H}_A^{(i)}$ é $n - (u + q) = n - u - 1 = n - \text{car}(\mathbf{A} \ \mathbf{c}_i)$.

Como $\hat{\mathbf{x}}_0$ (2.29) perde sua propriedade de melhor estimador linear imparcial (BLUE) em $\mathcal{H}_A^{(i)}$, o BLUE de \mathbf{x} em $\mathcal{H}_A^{(i)}$ se torna:

$$\hat{\mathbf{x}}_i = \bar{\mathbf{A}}_i^+ \mathbf{y}, \forall i = 1, \dots, n \quad (2.32)$$

com $\bar{\mathbf{A}}_i = \begin{pmatrix} \mathbf{A} & \mathbf{c}_i \end{pmatrix}$ sendo a matriz \mathbf{A} aumenta pelo vetor \mathbf{c}_i , e $\bar{\mathbf{A}}_i^+ = (\bar{\mathbf{A}}_i^T \mathbf{Q}_e^{-1} \bar{\mathbf{A}}_i)^{-1} \bar{\mathbf{A}}_i^T \mathbf{Q}_e^{-1}$.

Sob a hipótese nula \mathcal{H}_0 , temos a seguinte distribuição conjunta de \mathbf{x} e \mathbf{e} :

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{e} \end{bmatrix} \underset{\mathcal{H}_0}{\sim} \mathcal{N} \left(\begin{bmatrix} \mathbf{x} \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{Q}_{\hat{\mathbf{x}}_0} & 0 \\ 0 & \mathbf{Q}_{\hat{\mathbf{e}}_0} \end{bmatrix} \right) \quad (2.33)$$

As matrizes covariâncias, $\mathbf{Q}_{\hat{\mathbf{x}}_0}$ e $\mathbf{Q}_{\hat{\mathbf{e}}_0}$, podem ser obtidas como:

$$\begin{cases} \mathbf{Q}_{\hat{\mathbf{x}}_0} = (\mathbf{A}^T \mathbf{Q}_e^{-1} \mathbf{A})^{-1} \\ \mathbf{Q}_{\hat{\mathbf{e}}_0} = \mathbf{P}_A^\perp \mathbf{Q}_e \mathbf{P}_A^{\perp T} \end{cases} \quad (2.34)$$

com $\mathbf{P}_A = \mathbf{A}(\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W}$ e $\mathbf{P}_A^\perp = \mathbf{I} - \mathbf{P}_A$ sendo dois projetores ortogonais.

Como a esperança (média) do vetor dos resíduos estimados ($\hat{\mathbf{e}}_0$) é zero e estocasticamente independente de $\hat{\mathbf{x}}_0$, ele contém todas as informações úteis disponíveis para testar a validade de \mathcal{H}_0 .

Considerando que uma dada hipótese alternativa, digamos $\mathcal{H}_A^{(j)}$, seja detectada. Neste caso, a distribuição associada a ela é dada como:

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{e} \end{bmatrix} \underset{\mathcal{H}_j}{\sim} \mathcal{N} \left(\begin{bmatrix} \mathbf{I}_u & \mathbf{A}^+ \mathbf{c}_j \\ 0 & \mathbf{R} \mathbf{c}_j \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \nabla_j \end{bmatrix}, \begin{bmatrix} \mathbf{Q}_{\hat{\mathbf{x}}_0} & 0 \\ 0 & \mathbf{Q}_{\hat{\mathbf{e}}_0} \end{bmatrix} \right) \quad (2.35)$$

com a matriz $\mathbf{I}_u \in \mathbb{R}^{u \times u}$ sendo identidade.

De acordo com (2.35), observa-se que $\mathbf{c}_j \nabla_j$ é propagado de forma diferente nas médias (esperanças) de $\hat{\mathbf{x}}_0$ e $\hat{\mathbf{e}}_0$. Por essa razão, podemos subdividir o efeito do outlier em

três categorias, como segue:

$$\begin{cases} \nabla_{\mathbf{y}} = \mathbf{c}_j \nabla_j \text{ efeito do outlier nas observações} \\ \nabla_{\hat{\mathbf{x}}_0} = \mathbf{A}^+ \nabla_{\mathbf{y}} \text{ efeito do outlier nos parâmetros} \\ \nabla_{\hat{\mathbf{e}}_0} = \mathbf{R} \nabla_{\mathbf{y}} \text{ efeito do outlier nos resíduos} \end{cases} \quad (2.36)$$

Os efeitos acima (2.36) estão relacionados da seguinte forma:

$$\begin{cases} \nabla_{\mathbf{y}} = P_A \nabla_{\mathbf{y}} + P_A^\perp \nabla_{\mathbf{y}} = \mathbf{U} \nabla_{\mathbf{y}} + \mathbf{R} \nabla_{\mathbf{y}} \\ \nabla_{\mathbf{y}} = \mathbf{A} \nabla_{\hat{\mathbf{x}}_0} + \nabla_{\hat{\mathbf{e}}_0} \end{cases} \quad (2.37)$$

com $P_A = \mathbf{A} \mathbf{A}^+ = \mathbf{U}$ e $P_A^\perp = \mathbf{I}_n - P_A = \mathbf{R}$ sendo os projetores ortogonais. Essa decomposição das influências de um outlier pode ser visto com detalhes em Klein (2011, tópico 4.1 Números de Redundância e de Absorção, página 102).

A razão tendência-ruído (*bias-to-noise ratio* — BNR) de $\nabla_{\mathbf{y}}$, $\nabla_{\hat{\mathbf{x}}_0}$ e $\nabla_{\hat{\mathbf{e}}_0}$ é dada respectivamente por:

$$b_{\mathbf{y}} = \|\nabla_{\mathbf{y}}\|_{Q_e}, \quad b_{\hat{\mathbf{x}}_0} = \|\nabla_{\hat{\mathbf{x}}_0}\|_{Q_{\hat{\mathbf{x}}_0}} \quad e \quad b_{\hat{\mathbf{e}}_0} = \|\nabla_{\hat{\mathbf{e}}_0}\|_{Q_{\hat{\mathbf{e}}_0}} \quad (2.38)$$

Nota-se que os BNRs acima (2.38) representam a soma do quadrado de seus elementos $\nabla_{(y, \hat{x}_0, \hat{e}_0)}$ ponderados pelas suas respectivas matrizes de covariâncias. O BNR expressa a relação entre a parte associada com a tendência (outlier) e a parte aleatória (incerteza). Mais detalhes ver Klein (2011).

Os BNRs podem ser geometricamente relacionados por meio da relação de Pitágoras, ou seja,

$$b_{\mathbf{y}}^2 = b_{\hat{\mathbf{x}}_0}^2 + b_{\hat{\mathbf{e}}_0}^2 \quad (2.39)$$

A Figura 6 mostra o ângulo θ entre $\nabla_{\mathbf{y}}$ e $P_A \nabla_{\mathbf{y}}$. Ele indica a relação entre o efeito do outlier nos resíduos e nos parâmetros do modelo. Nota-se que este ângulo θ também determina a razão entre o BNR associado ao efeito do outlier nos resíduos ($b_{\hat{\mathbf{e}}_0}$) e o BNR associado com o efeito do outlier nos parâmetros ($b_{\hat{\mathbf{x}}_0}$), ou seja:

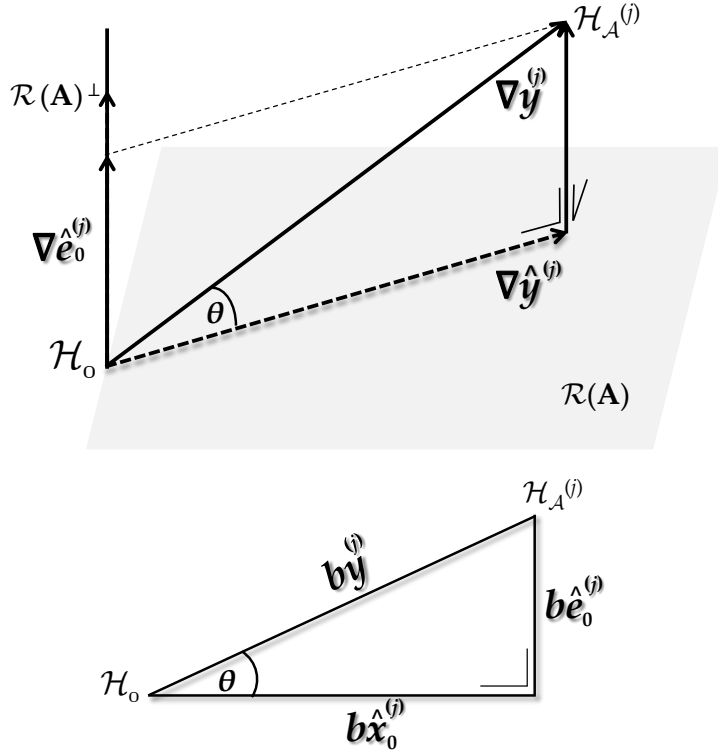
$$b_{\hat{\mathbf{e}}_0} = b_{\hat{\mathbf{x}}_0} \tan(\theta) \quad (2.40)$$

Nota-se pela Figura (6) que quanto menor o ângulo θ , maior é o alinhamento entre $\nabla_{\mathbf{y}}$ e $\mathcal{R}(\mathbf{A})$ e, portanto, maior é o efeito de um outlier de um possível outlier nos parâmetros. O ângulo θ por si só é determinado pela matriz que define o modelo de outlier \mathbf{C} em (2.3) e pelo modelo funcional (\mathbf{A}) e estocástico (Q_e).

Podemos analisar também o efeito da adaptação quando uma hipótese alternativa $\mathcal{H}_A^{(j)}$ é detectada. Os estimadores lineares imparciais (BLUEs) de \mathbf{x} e ∇_j sob a hipótese alternativa $\mathcal{H}_A^{(j)}$ são dados respectivamente por:

$$\hat{\mathbf{x}}_j = \bar{\mathbf{A}}_j^+ \mathbf{y} = \hat{\mathbf{x}}_0 - \mathbf{A}^+ \mathbf{c}_j \hat{\nabla}_j \quad (2.41)$$

Figura 6 – Decomposição dos efeitos de um outlier no espaço das observações, parâmetros e resíduos.



Fonte: Produzido pelos autores.

$$\hat{\nabla}_j = \mathbf{c}_j^T \mathbf{R} \mathbf{y} \quad (2.42)$$

O resultado em (2.41) mostra como $\hat{\mathbf{x}}_0$ deve ser adaptado ao mudar do BLUE de \mathcal{H}_0 para o BLUE de $\mathcal{H}_A^{(j)}$. Essa adaptação mostra o efeito da remoção de um outlier localizado na j -ésima sobre o estimador e, conseqüentemente, sobre os parâmetros do modelo.

Podemos usar a relação dada em (2.37) e analisar a relação entre os estimadores BLUEs de diferentes hipóteses.

Vamos considerar agora a relação dada por (2.37). Nesse caso, podemos decompor o vetor de incertezas em dois componentes: um componente associado apenas aos efeitos aleatórios (\mathbf{e}) e outro componente associado com o outlier (∇_i). Desta forma, temos para cada hipótese $i = 0, \dots, n$:

$$\boldsymbol{\varepsilon}_i = \mathbf{e}_i + \mathbf{c}_i \nabla_i, \forall i = 0, \dots, n \quad (2.43)$$

Sob a hipótese nula \mathcal{H}_0 , $\mathbf{c}_i \nabla_i$ é nulo e, portanto, $\mathbf{e}_0 = \boldsymbol{\varepsilon}_0$.

O estimador de $\boldsymbol{\varepsilon}_i$ é dado por:

$$\hat{\boldsymbol{\varepsilon}}_i = \mathbf{R}(\mathbf{e}_i + \mathbf{c}_i \nabla_i) \quad (2.44)$$

Sobre a hipótese nula \mathcal{H}_0 , o vetor de incertezas estimado ($\hat{\boldsymbol{\varepsilon}}_0$) é função apenas dos efeitos aleatórios, ou seja, $\hat{\boldsymbol{\varepsilon}}_0 = \hat{\boldsymbol{\varepsilon}}_0 = \mathbf{R}\boldsymbol{\varepsilon}_0$.

Com base na relação dada em (2.44), podemos estabelecer a seguinte transformação entre os BLUEs de \boldsymbol{x} e $\boldsymbol{\varepsilon}$ sob as $n + 1$ hipóteses \mathcal{H}_i para $i = 0, \dots, n$:

$$\begin{bmatrix} \hat{\boldsymbol{x}}_i \\ \hat{\boldsymbol{\varepsilon}}_i \end{bmatrix} = \begin{bmatrix} \mathbf{I}_u & -T_i \\ 0 & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{x}}_0 \\ \hat{\boldsymbol{\varepsilon}}_0 \end{bmatrix} T_i = \begin{cases} 0 & \text{para } i = 0 \\ \mathbf{A}^+ \mathbf{c}_i \mathbf{c}_i^T \mathbf{R} & \text{para } i \neq 0 \end{cases} \quad (2.45)$$

com $\mathbf{I}_u \in \mathbb{R}^{u \times u}$ e $\mathbf{I}_n \in \mathbb{R}^{n \times n}$ sendo as matrizes identidades associadas ao espaço dos parâmetros e observações, respectivamente. Observem que esta transformação é da forma bloco-triangular, e sua inversa pode ser obtida simplesmente invertendo o sinal de $-T_i$ para T_i . Essa transformação é conhecida como *Tienstra-Transformation* (TIENSTRA, 1956).

Notem que quando $i = 0$, estamos sob \mathcal{H}_0 . Neste caso, os estimadores de \boldsymbol{x} e $\boldsymbol{\varepsilon}$ têm suas respectivas distribuições dadas por (2.3).

A distribuição de (2.45) sob uma hipótese alternativa, digamos (\mathcal{H}_a) com $\mathbf{c}_a \nabla_a$, é dada por:

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{\varepsilon} \end{bmatrix} \underset{\mathcal{H}_a}{\sim} \mathcal{N} \left(\begin{bmatrix} \boldsymbol{x} + \mathbf{A}^+ \mathbf{c}_a \mathbf{c}_a^T \mathbf{R} \mathbf{c}_a \nabla_a \\ \mathbf{R} \mathbf{c}_a \nabla_a \end{bmatrix}, \begin{bmatrix} \mathbf{Q}_{\hat{\boldsymbol{x}}_0} + T_a \mathbf{Q}_{\hat{\boldsymbol{\varepsilon}}_a} T_a^T & -T_i \mathbf{Q}_{\hat{\boldsymbol{\varepsilon}}_a} \\ -\mathbf{Q}_{\hat{\boldsymbol{\varepsilon}}_a} T_a^T & \mathbf{Q}_{\hat{\boldsymbol{\varepsilon}}_a} \end{bmatrix} \right) \quad (2.46)$$

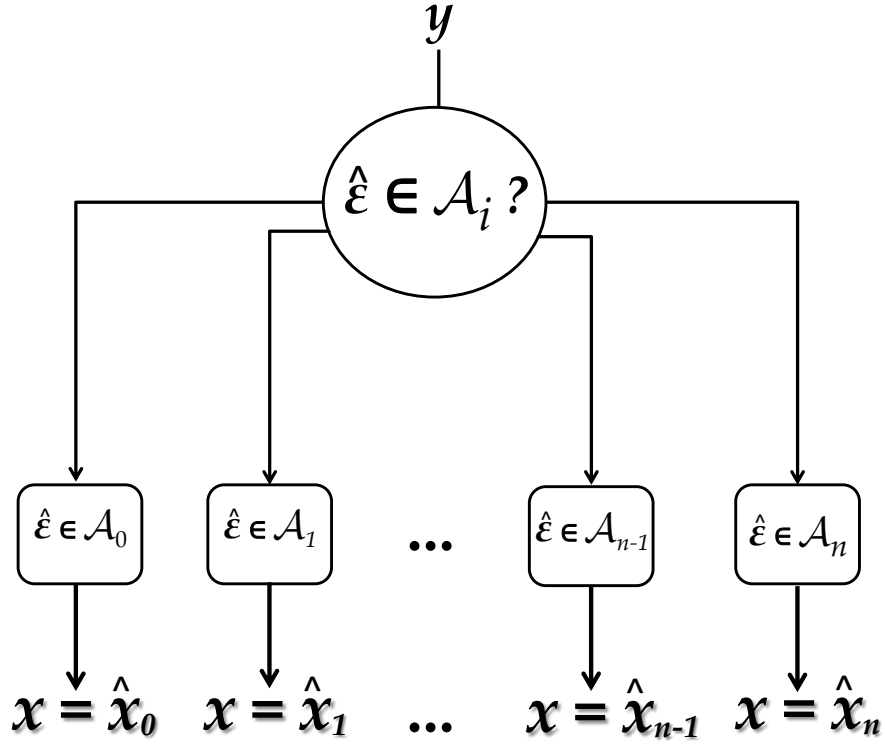
A distribuição em (2.46) foi obtida por meio da aplicação dos operadores de esperança matemática $\mathbb{E}\{\cdot\}$ e dispersão $\mathbb{D}\{\cdot\}$ em (2.45). Essa adaptação foi dada para o caso no qual apenas uma única hipótese alternativa está presente. Nota-se que $\mathcal{H}_A^{(a)}$ foi fixada. Logo, qualquer outro estimador sob $\mathcal{H}_A^{(a)}$ será tendencioso, ou seja, $\mathbb{E}\{\boldsymbol{x}_{i \neq a} - \boldsymbol{x} | \mathcal{H}_a\} = \mathbf{A}^+ \mathbf{c}_{i \neq a} \mathbf{c}_{i \neq a}^T \mathbf{R} \mathbf{c}_{i \neq a} \nabla_a$. Como já comentado, o DSI envolve múltiplas hipóteses com adaptações iterativas e não apenas uma única hipótese alternativa.

A transformação dada em (2.45) mostra como o vetor dos resíduos $\hat{\boldsymbol{\varepsilon}}$ desempenha o seu papel na conexão dos BLUEs das diferentes hipóteses. Essa relação, no entanto, ainda não incorpora o resultado do teste. Em outras palavras, as probabilidades associadas com o procedimento de teste não são levadas em consideração nas estimativas dos parâmetros do modelo. Para fazer isso, deve-se particionar o espaço amostral dos resíduos de acordo com o número de hipóteses presentes, i.e. $\cup_{i=0}^n \mathcal{A}_i = \mathbb{R}^n$ e $\mathcal{A}_i \cap \mathcal{A}_j = \{0\}$ $i \neq j$, em que \mathcal{A} representa o espaço amostral dos resíduos. Então, o procedimento de teste é definido de modo que \mathcal{H}_i seja selecionada se e somente se $\hat{\boldsymbol{\varepsilon}}_i \in \mathcal{A}_i$. A Figura (7) ilustra o efeito de considerar um procedimento de teste nas estimativas dos parâmetros de interesse.

No caso do DSI, a hipótese nula \mathcal{H}_0 é aceita quando $\hat{\boldsymbol{\varepsilon}}_i \in \mathcal{A}_0$, com a região de aceitação \mathcal{A}_0 dada por:

$$\mathcal{A}_0 = \{\hat{\boldsymbol{\varepsilon}}_i \in \mathbb{R}^n \mid |\boldsymbol{w}_i| \ (\forall i = 1, \dots, n) \leq \hat{k} \in \mathbb{R}^+\} \quad (2.47)$$

Figura 7 – Resultado de $\hat{\mathbf{x}}$ em função do procedimento de teste.



Fonte: Produzido pelos autores.

O complemento de \mathcal{A}_0 , ou seja a região crítica $\mathcal{A}_{i \neq 0}$, é dado por:

$$\mathcal{A}_{i \neq 0} = \{\hat{\boldsymbol{\varepsilon}}_i \in \mathbb{R}^n \mid |\mathbf{w}_i| \ (\forall i = 1, \dots, n) > \hat{k} \in \mathbb{R}^+\} \quad (2.48)$$

As expressões (2.47) e (2.48) estão associadas com a detecção de outlier, pois somente informa se a hipótese nula deve ser aceita ou rejeitada. Se a hipótese nula for rejeitada, devemos encontrar a hipótese alternativa que levou à rejeição da hipótese nula, ou seja, estamos interessados em identificar o outlier. A identificação ocorre quando:

$$\mathcal{A}_{i=i} = \{\hat{\boldsymbol{\varepsilon}}_i \in \mathbb{R}^n \mid \max_{i \in \{1, \dots, n\}} |\mathbf{w}_i| > \hat{k} \in \mathbb{R}^+\} \quad (2.49)$$

O tamanho de \mathcal{A}_0 é determinado por meio da taxa de erro da família de testes α' (FWER – *family-wise error rate*), ou seja, $\mathcal{P}\{\hat{\boldsymbol{\varepsilon}}_i \in \mathcal{A}_0 | \mathcal{H}_0\} = 1 - \alpha'$.

O FWER controla o Erro Tipo I para o caso de múltiplas hipóteses, ou seja, ele garante que a probabilidade de rejeitar erroneamente ao menos uma hipótese nula não aumente com o número de testes envolvidos. Uma forma aproximada de controlar o FWER foi dada em (2.15) com o valor crítico \hat{k} calculado por (2.16). Aquele procedimento, entretanto, despreza a inevitável dependência entre as hipóteses envolvidas. A dependência se deve à natureza da matriz covariância dos resíduos $\mathbf{Q}_{\hat{\boldsymbol{\varepsilon}}}$ de nunca ser bloco-diagonal. Para superar esse problema e de forma similar a [Lehmann \(2012\)](#), tratamos o *max-w*

como sendo a estatística de teste do DSI e, portanto, ela possui sua própria distribuição de probabilidade. No Capítulo 3 fornecemos um método baseado em Monte Carlo para derivar a f.d.p. de max- w de modo que usuário possa controlar efetivamente o Erro Tipo I. Importante destacar que na falta de um procedimento de teste, temos $\mathcal{A}_0 = \mathbb{R}$.

De acordo com a Figura (6), pode-se escrever:

$$\bar{\mathbf{x}}_{DSI} = \begin{cases} \hat{\mathbf{x}}_0 & \text{se } \hat{\boldsymbol{\varepsilon}}_0 \in \mathcal{A}_0 \\ \hat{\mathbf{x}}_i & \text{se } \hat{\boldsymbol{\varepsilon}}_i \in \mathcal{A}_{i \neq 0} \quad \forall i = 1, \dots, n \end{cases} \quad (2.50)$$

Utilizando as funções indicadoras $f_i(\hat{\boldsymbol{\varepsilon}}_i)$ das regiões \mathcal{A}_i , o estimador DSI ($\bar{\mathbf{x}}_{DSI}$) pode ser escrito na forma compacta como:

$$\bar{\mathbf{x}}_{DSI} = \sum_{i=0}^n \hat{\mathbf{x}}_i f_i(\hat{\boldsymbol{\varepsilon}}_i) \quad (2.51)$$

$$f_i(\hat{\boldsymbol{\varepsilon}}_i) = \begin{cases} 1 & \text{para } \hat{\boldsymbol{\varepsilon}}_i \in \mathcal{A}_i \\ 0 & \text{para } \hat{\boldsymbol{\varepsilon}}_i \notin \mathcal{A}_i \end{cases}$$

A relação dada em (2.51) comprova que o DSI não é somente teste, mas uma combinação de estimação e teste e, portanto, os resultados finais também irão depender dos níveis de probabilidade associados com o DSI. Importante ressaltar que alterando o valor do FWER, pode-se alterar o estimador $\bar{\mathbf{x}}_{DSI}$ daquela amostra, ou seja, o estimador depende do resultado do teste.

Se fizermos o uso da transformação dada em (2.45), então é possível escrever a relação entre o estimador $\bar{\mathbf{x}}_{DSI}$ e $\hat{\boldsymbol{\varepsilon}}_i$ como:

$$\begin{bmatrix} \bar{\mathbf{x}}_{DSI} \\ \hat{\boldsymbol{\varepsilon}}_i \end{bmatrix} = \begin{bmatrix} \mathbf{I}_u & -\bar{T}(\hat{\boldsymbol{\varepsilon}}_i) \\ 0 & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_0 \\ \hat{\boldsymbol{\varepsilon}}_i \end{bmatrix} \quad (2.52)$$

com

$$\bar{T}(\hat{\boldsymbol{\varepsilon}}_i) = \sum_{i=0}^n T_i f_i(\hat{\boldsymbol{\varepsilon}}_i) \quad (2.53)$$

Embora a estrutura em (2.52) seja semelhante à transformação linear em (2.45), a transformação em (2.52) não é linear devido à presença de $\hat{\boldsymbol{\varepsilon}}_i$ em $\bar{T}(\boldsymbol{\varepsilon}_i)$. Isso implica que o estimador DSI não terá distribuição normal, mesmo que todos os estimadores individuais $\hat{\mathbf{x}}_i$, $i = 0, \dots, n$ sejam normalmente distribuídos.

Observem que as incertezas das decisões do DSI (níveis de probabilidade de correta detecção e correta identificação) são canalizadas por meio de $f_i(\boldsymbol{\varepsilon}_i)$, enquanto que a incerteza da estimação é canalizada por $\hat{\mathbf{x}}_i$. Portanto, o DSI é um estimador de \mathbf{x} .

Qualquer tipo de adaptação do modelo, como a remoção de outlier, produzirá um estimador tendencioso sob a hipótese alternativa $\mathcal{H}_A^{(i)}$ (TEUNISSEN, 2018). Porém, a

tendência produzida após a adaptação será sempre menor que a tendência originalmente produzida sem aplicação de um teste estatístico prosseguido por uma adaptação do modelo. Em alguns casos, é possível até mesmo remover a tendência, fazendo com que o DSI mantenha a propriedade de BLUE sob uma $\mathcal{H}_A^{(i)}$ (ZAMINPARDAZ; TEUNISSEN, 2019b).

As decisões do DSI dependem do vetor dos resíduos $\hat{\epsilon}_i$ e de sua respectiva matriz de covariância $\mathbf{Q}_{\hat{\epsilon}_i}$. Na ausência de um outlier, ou seja quando a hipótese nula é verdadeira \mathcal{H}_0 , descrevemos a probabilidade de cometer o Erro Tipo I e uma forma aproximada de controlar o FWER. Porém, já destacamos que essa forma não considera a dependência entre as hipóteses. No cenário com a presença de um outlier nos dados observados, ou seja sob $\mathcal{H}_A^{(i)}$, as probabilidades do *Data-Snooping* para múltiplas hipóteses foram descritas anteriormente, especificamente na Tabela 1. Entretanto, como visto, o DSI é um procedimento iterativo de eliminação de outlier que envolve múltiplas hipóteses com adaptações.

Na presença de um outlier, ou seja quando $\mathcal{H}_A^{(i)}$ é verdadeira, os níveis de probabilidade do DSI são decompostas nas seguintes classes:

- \mathcal{P}_{CI} (*correct identification*): probabilidade de identificar e remover corretamente o outlier.
- \mathcal{P}_{MD} (*missed detection*): probabilidade de não detectar, ou seja, probabilidade de cometer falso negativo (Erro Tipo II).
- \mathcal{P}_{WE} (*wrong exclusion*): probabilidade de identificar e remover uma observação, enquanto o outlier permanece nos dados, ou seja, é a probabilidade de cometer o Erro Tipo III.
- \mathcal{P}_{over+} (*over-identification positive*): probabilidade de identificar e remover corretamente o outlier e outras observações;
- \mathcal{P}_{over-} (*over-identification negative*): probabilidade de identificar e remover mais de uma observação como sendo outlier, enquanto o outlier permanece nos dados observados.
- \mathcal{P}_{ol} (*statistical overlap*): probabilidade de detectar simultaneamente duas (ou mais) observações como outliers. Isso ocorre nos casos em que uma hipótese alternativa tem a mesma distribuição que outra. Em outras palavras, essas hipóteses não podem ser distinguidas; ou seja, elas são inseparáveis e não é possível identificar um outlier, embora a hipótese nula tenha sido corretamente rejeitada.

Infelizmente, as f.d.p.'s associadas com o procedimento DSI possuem integrais multivariadas em regiões complexas. Como um modelo analítico de fácil solução é desconhecido para o DSI, é necessário recorrer as simulações computacionais, por exemplo, o método de

Monte Carlo (MMC). Diferentemente da época de Baarda, hoje, temos a nossa disposição computadores potentes o suficiente para estudar, descrever e entender certos problemas que são analiticamente intratáveis. Dentro deste contexto, o MMC surge como uma ferramenta essencial para tratar de processos que dependem de fatores aleatórios, como é o caso do DSI.

2.4 Método de Monte Carlo no contexto da Geodésia

Intitula-se por método de Monte Carlo (MMC) qualquer procedimento que se baseia em amostragens aleatórias massivas para obter resultados numéricos (HROMKOVIČ, 2001). O MMC é usado para solucionar problemas que envolvem um grande número de variáveis, vários tipos de distribuições estatísticas e modelos (lineares e não-lineares) complexos (ROBERT; CASELLA, 2004). O MMC tem sido considerado uma ferramenta fundamental em muitas áreas da ciência moderna, com aplicações diversas. Pode-se citar, por exemplo, estudos sobre o fluxo de veículos no trânsito, econometria, evolução estelar, exploração de poços de petróleo, estudos de dosimetria para tratamento de câncer, projeto de reatores nucleares e de aceleradores de partículas, entre outros (SANTOS et al., 2014; WIEGAND, 2006; DU; GENG; SUN, 2009; CHATTERJEE et al., 2010; KOK; KAYA; AKIN, 2006; KIM et al., 2016).

O uso de amostragens aleatórias para resolver problemas matemáticos parece ter sido documentado pela primeira vez por Georges Louis Leclerc, Conde de Buffon, em 1772. Desde então, ao longo da história, diversos experimentos baseados em amostragens aleatórias foram conduzidos, como exemplo (STIGLER, 2002):

- Em 1786, Pierre S. Laplace obteve uma aproximação do número π (pi) por amostragem aleatória;
- Em 1899, John W. S. Rayleigh mostrou que um passeio aleatório (*random walk*), poderia fornecer uma solução aproximada para uma equação diferencial parabólica;
- Em 1901, Kelvin parece ter usado amostragem aleatória para avaliar integrais que aparecem na teoria cinética dos gases;
- Em 1908, William S. Gosset, *mestre cervejeiro*, publicou com o pseudônimo *Student* artigos referentes à estimativa dos coeficientes de correlação em sua distribuição t (conhecida como t-de-Student) por meio de amostragens aleatórias;
- Em 1931, Andrey N. Kolmogorov mostrou a relação entre os processos estocásticos de Markov e certas equações diferenciais.

Entretanto, o grande impulso no desenvolvimento do MMC deve-se aos estudos conduzidos por Enrico Fermi, Stanislaw Ulam, John von Neumann e Nicholas Metropolis para solucionar problemas de difusão de nêutrons no desenvolvimento da bomba atômica em 1944. No final da década de 1940 e início da década de 1950 surgiram trabalhos descrevendo o MMC e sua aplicação para resolver problemas de transporte de partículas e radiação. O nome “Monte Carlo” é creditado a Nicholas Metropolis e refere-se à cidade europeia de mesmo nome, famosa pelos seus cassinos e jogos de azar, como a roleta, a qual em última análise é um gerador de números aleatórios (METROPOLIS, 1987).

Apesar dos avanços da ciência e da tecnologia para resolver sistemas altamente complexos, uma das principais desvantagens do MMC até a década de 1980 residia no tempo de sua execução. Os recursos computacionais naquela época limitavam o uso do MMC. Hoje, o cenário computacional é diferente. A disponibilidade de microcomputadores de baixo custo e alto desempenho tem impulsionado fortemente aplicações do MMC em geodésia, em especial a área de controle de qualidade (Wyszkowska, 2017; NIEMEIER; TENGEN, 2017; ALKHATIB; NEUMANN; KUTTERER, 2009; KOCH, 2008; LEHMANN; SCHEFFLER, 2011; LEHMANN, 2015; LEHMANN, 2013b; LEHMANN; LÖSLER, 2017; LEHMANN, 2012; SURACI; OLIVEIRA, 2010; ROFATTO; MATSUOKA; KLEIN, 2017; ROFATTO; MATSUOKA; KLEIN, 2018; KLEIN et al., 2015; KLEIN et al., 2019; IMPARATO; TEUNISSEN; TIBERIUS, 2019; ROFATTO et al., 2020a).

As etapas básicas de um procedimento baseado em MMC incluem:

1. Definir o problema por meio de equações , ou seja, deve-se identificar o modelo matemático com todas as suas variáveis. Por exemplo, a Equação (2.44) descreve a relação do vetor dos resíduos em função dos vetores de incertezas \mathbf{e}_i e do modelo de outlier $\mathbf{c}_i \nabla_i$.
2. Definir as f.d.p.'s das variáveis de entrada. Por exemplo, algumas variáveis podem seguir a distribuição normal, enquanto outras seguem distribuição triangular ou uniforme. Tomando novamente a Equação (2.44) como exemplo, podemos considerar $\mathbf{e} \sim \mathcal{N}(0, \mathbf{Q}_e)$.
3. Gerar sinteticamente as variáveis aleatórias com base em suas distribuições definidas na etapa (2). Para isso, um gerador de números (pseudo)aleatórios confiável deve estar disponível. Neste etapa, deve-se definir o número de experimentos Monte Carlo (denotado por m). O número de experimentos Monte Carlo define o número de vezes que cada variável será artificialmente gerada. Por exemplo, se $m = 1000$ e $n = 10$, então teríamos $\boldsymbol{\varepsilon}_i^{(1)}, \boldsymbol{\varepsilon}_i^{(2)}, \dots, \boldsymbol{\varepsilon}_i^{(m)} \forall i = 1, \dots, 10$ vetores de incerteza sinteticamente gerados, onde o subscrito é a dimensão do vetor de acordo com o número de observáveis disponíveis (n) e o sobrescrito é o número de experimentos

Monte Carlo (m). Isso significa que para cada i observável, teríamos m valores disponíveis.

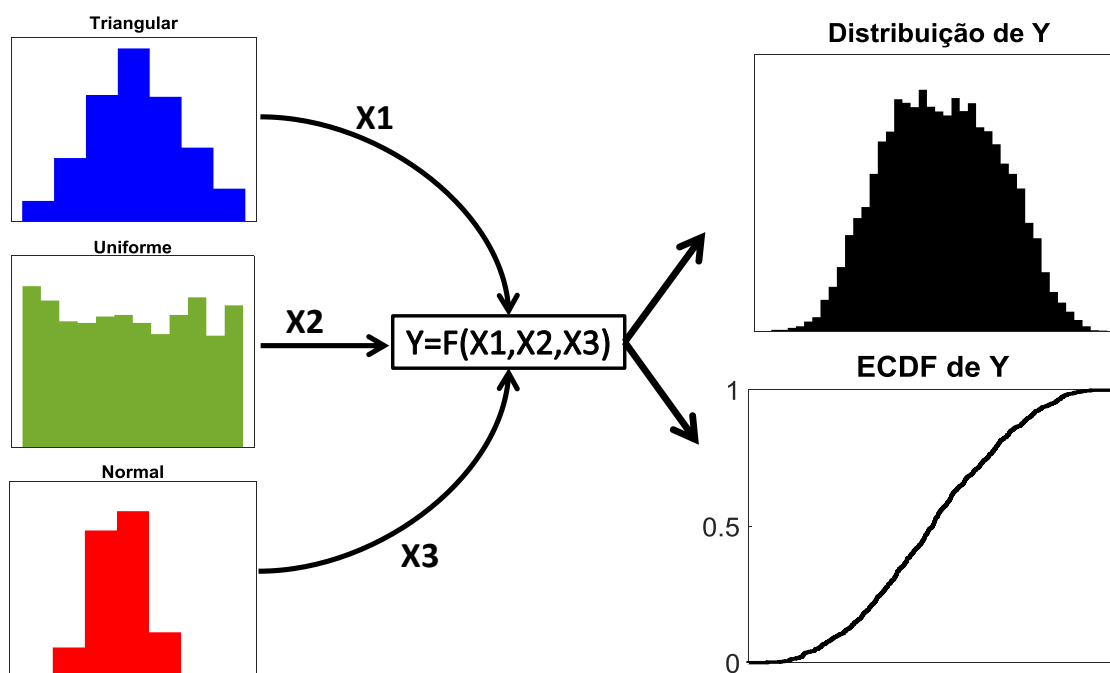
4. As amostras geradas são inseridas no modelo definido na Etapa 1. Tomando o exemplo anterior, se $m = 1000$, teríamos 1000 modelos gerados sinteticamente, o que permite uma série de análises. Podemos, por exemplo:
 - Classificar os m modelos em ordem crescente com o objetivo de obter a função de distribuição acumulada empírica (f.d.a.e.), também conhecida como ECDF (*Empirical Cumulative Density Function*).
 - Calcular o valor mais provável e a incerteza de uma variável aleatória de interesse.
 - Construir intervalos de confiança.
 - Calcular valores extremos, ou críticos.
 - Estimar a probabilidade de um certo evento, dividindo o número de ocorrência do evento pelo número total de experimentos m . Esta última aplicação em especial é utilizada nessa Tese. Mais detalhes são fornecidos no Capítulo 3 e 4.

A Figura (8) ilustra um exemplo de aplicação do MMC em um problema de propagação de distribuições. A Figura (9) ilustra um outro exemplo no qual o MMC é usado para descrever o modelo em (2.44) sobre a hipótese nula \mathcal{H}_0 e sobre a hipótese alternativa. \mathcal{H}_a [azul — resíduos calculados com base em $\hat{\boldsymbol{\epsilon}}_0 = \hat{\boldsymbol{\epsilon}}_0 = \mathbf{R}\mathbf{e}$; vermelho — resíduos calculados com base em $\hat{\boldsymbol{\epsilon}}_a = \mathbf{R}(\mathbf{e} + \mathbf{c}_a \nabla_a)$].

Diferentemente de métodos determinísticos, o MMC não resolve uma equação de forma explícita, como pode ser observado nos exemplos das Figuras (8) e (9). Todas as aplicações do MMC baseiam-se na construção de um modelo probabilístico para descrever o(s) fenômeno(s) de interesse.

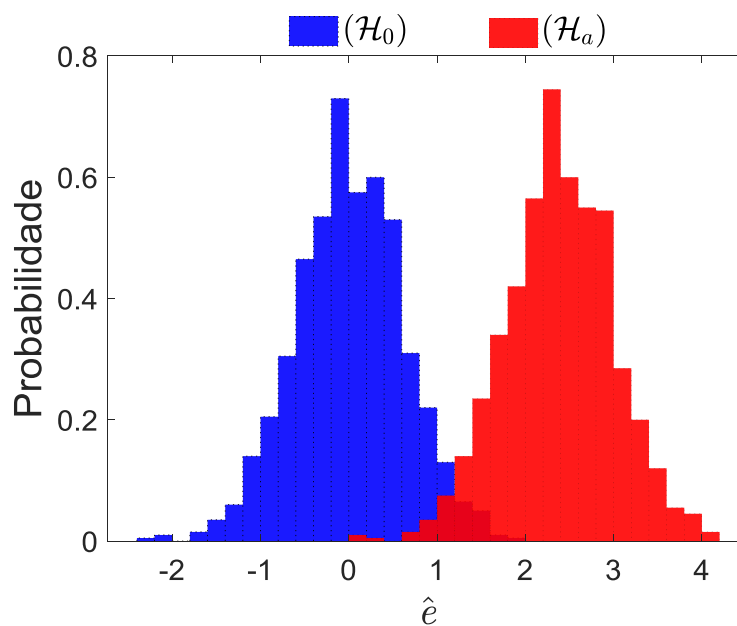
O ponto chave do MMC são os geradores de números aleatórios (GNA). As variáveis aleatórias de interesse são sinteticamente produzidas com auxílio de um GNA. Os GNA são úteis para diversas finalidades, como gerar chaves de criptografia de dados, simular e modelar fenômenos complexos e selecionar amostras aleatórias de conjuntos de dados maiores. Em termos de computação inteligente, os GNAs são usados na inicialização das matrizes de peso em redes neurais artificiais; os algoritmos meta-heurísticos empregam os GNAs para encontrar soluções ótimas para problemas reconhecidamente difíceis de se solucionarem por métodos analíticos, tais como algoritmos genéticos, sistemas imunológicos artificiais, algoritmos de enxames (abelhas, formigas, morcegos, baratas, etc.), inteligência coletiva, sistemas nebulosos (fuzzy), entre outros. Alguns exemplos são Koch (2017), Baselga (2011), Koch et al. (2019), Jwo e Chen (2006).

Figura 8 – Ilustração da aplicação do Método de Monte Carlo na propagação de distribuições de probabilidade.



Fonte: Produzido pelos autores.

Figura 9 – Exemplo de aplicação do Método Monte Carlo na descrição de modelos sob diferentes hipóteses.



Fonte: Produzido pelos autores.

Duas formas distintas são normalmente empregadas para gerar tais sequências de números aleatórios: *Hardware-Random Number Generator* (HRNG) e *Pseudo-Random Number Generator* (PRNG). O HRNG é um dispositivo que gera números aleatórios a partir de ruídos (flutuações probabilísticas) advindos de processos físicos, tais como ruídos atmosféricos (<https://www.random.org/>), ruído eletrônico gerado pela agitação térmica dos portadores de carga (geralmente os elétrons) dentro de um condutor elétrico em equilíbrio, efeito fotoelétrico, entre outros fenômenos quânticos. Esses processos estocásticos são, em teoria, completamente imprevisíveis, e as afirmações de imprevisibilidade da teoria estão sujeitas a testes experimentais. Um PRNG, por outro lado, não é um dispositivo, mas um algoritmo que gera números cujas propriedades se aproximam das propriedades de sequências de números aleatórios. Por esta razão, são chamados de Geradores de Números Pseudo-Aleatórios (GNPA). Embora as sequências mais próximas de serem realmente aleatórias possam ser geradas usando um HRNG, os GNPA são importantes na prática devido à sua velocidade e a possibilidade de reproduzir a sequência de números gerados de um determinado experimento.

A sequência gerada por um GNPA não é verdadeiramente aleatória, pois é determinada por um valor inicial, chamado semente (*seed*). Embora a semente inicial possa ser obtida a partir de uma fonte de ruído aleatória, o algoritmo empregado para geração da sequência de números aleatórios é puramente determinístico. Se a semente inicial é conhecida, é possível reproduzir a sequência. Um algoritmo tem limitações neste processo, pois em um dado momento a semente inicial deverá ser reutilizada e, portanto, a sequência de números aleatórios será repetida. O comprimento antes de uma sequência de números se repetir é conhecido como período do GNPA. O período é estritamente limitado pelo comprimento da semente inicial, ou seja, pelo seu número de dígitos. Por exemplo, uma semente inicial com dois dígitos pode produzir no máximo 100 números aleatórios antes de reutilizar a semente inicial. Um GNPA de alta qualidade, tanto em termos de desempenho computacional quanto estatístico, é conhecido como Mersenne Twister (MATSUMOTO; NISHIMURA, 1998). Uma versão do Mersenne Twister disponível em muitas linguagens de programação, MT19937, tem um período de $2^{19937} - 1$. Devido às suas características, o Mersenne Twister (MT19937) é utilizado neste trabalho para gerar as sequências de números uniformemente distribuídos no intervalo de 0 a 1, ou seja, $\mathcal{U}[0, 1]$. Geralmente em geodésia e muitas outras ciências, entretanto, as observáveis seguem distribuição normal. Aqui, a transformação dada por Box e Muller (1958) é empregada para converter uma sequência de números gerados a partir da distribuição uniforme $\mathcal{U}[0, 1]$ para uma distribuição normal multivariada $\mathcal{N}(0, \mathbf{Q}_e)$. Essa transformação é conhecida como Box-Müller.

3 Material e Métodos

Aqui, apresentamos duas metodologias com objetivos distintos, porém complementares. A primeira se justifica pelo fato da estatística de teste do Data-Snooping possuir sua própria distribuição de probabilidade. Consequentemente, os valores críticos não estão disponíveis para o Data-Snooping. A fim de superar este problema e de forma similar à [Lehmann \(2012\)](#), recorreremos ao Método de Monte Carlo para construir a distribuição de probabilidade do teste envolvido no Data-Snooping, o que de certa forma viabilizou a obtenção do valor crítico em função do nível de significância estipulado. Quanto à segunda metodologia, trata-se de um procedimento computacional cujo objetivo foi estimar as probabilidades de sucesso e de falha do Data-Snooping, na presença de um outlier nos dados observados. Com base nos níveis de probabilidade do Data-Snooping, também foi possível obter os indicadores de sensibilidade para fins de detecção e identificação de outlier. É importante destacar que as estimativas das probabilidades do Data-Snooping são obtidas em função dos valores críticos obtidos pela metodologia anterior.

Importante destacar que método foi aprimorado ao longo do desenvolvimento desta pesquisa. Na sua primeira versão, os outliers eram sorteados uniformemente para um dado problema e, portanto, somente descrevia as probabilidades médias do problema, como pode ser visto em [Rofatto, Matsuoka e Klein \(2017\)](#). Já na segunda versão, os níveis de probabilidade associados com o DSI foram calculados para cada observação individual ([ROFATTO et al., 2020a](#)). Porém, naquela ocasião, o valor crítico foi calculado com base na distribuição normal, e nenhuma correção foi aplicada para controlar o Erro Tipo I. Na versão mais recente, os valores críticos associados com a estatística de teste do DSI foram otimizados por meio do Método de Monte Carlo. Isso significa que o usuário pode controlar efetivamente o Erro Tipo I. Além disso, propomos uma nova classe de probabilidade que não havia nas versões anteriores. Nós denominamos essa nova classe de sobreposição estatística (*statistical overlap* – \mathcal{P}_{ol}). Ela acontece sempre que o coeficiente de correlação entre as estatísticas de teste (2.19) é ± 1.00 (i.e.: $|\rho_{w_i, w_j}| = 1.00$). Nesse caso, o *max-w* (2.17) detecta duas ou mais hipóteses alternativas simultaneamente, pois as suas distribuições de probabilidade são idênticas. Por serem iguais, essas hipóteses não são distinguidas e, portanto, a identificação de um possível outlier não é possível.

Em contraste com *Data-Snooping* sem adaptação, a taxa de sucesso de detecção (\mathcal{P}_{CD}) do DSI depende da soma das probabilidades de correta identificação \mathcal{P}_{CI} , falsa exclusão \mathcal{P}_{WE} , sobre-identificação com a correta identificação \mathcal{P}_{over+} , sobre-identificação sem a correta identificação \mathcal{P}_{over-} e sobreposição estatística \mathcal{P}_{ol} , ou seja:

$$\mathcal{P}_{CD} = 1 - \mathcal{P}_{MD} = \mathcal{P}_{CI} + \mathcal{P}_{WE} + \mathcal{P}_{over+} + \mathcal{P}_{over-} + \mathcal{P}_{ol} \quad (3.1)$$

A probabilidade de correta detecção é o complemento da falha de detecção \mathcal{P}_{MD} , ou seja, $\mathcal{P}_{CD} = 1 - \mathcal{P}_{MD}$. Observem em (3.1) que a taxa de correta detecção \mathcal{P}_{CD} sempre estará disponível mesmo que a correta identificação seja nula $\mathcal{P}_{CI} = 0$. Entretanto, a identificação somente é possível se houver detecção, ou seja:

$$\mathcal{P}_{CI} = \mathcal{P}_{CD} - (\mathcal{P}_{WE} + \mathcal{P}_{over+} + \mathcal{P}_{over-} + \mathcal{P}_{ol}), \quad (3.2)$$

de modo que

$$\exists(\mathcal{P}_{CI}) \in [0, 1] \iff (\mathcal{P}_{CD}) > 0 \quad (3.3)$$

Importante mencionar novamente que a falsa exclusão, denotada por \mathcal{P}_{WE} , descreve a probabilidade de identificar e remover uma observação, enquanto o outlier permanece nos dados observados. Podemos decompor a falsa exclusão em função do número de vezes que cada observação foi erroneamente identificada e removida:

$$\mathcal{P}_{WE} = \sum_{i=1}^{n-1} \mathcal{P}_{WE(i)} = \sum_{i=1}^{n-1} \frac{n_{WE(i)}}{m} \quad (3.4)$$

Com base na Equação (3.4), podemos calcular um fator que pondera a contribuição de cada observação para a falsa exclusão (denotador por $p_{i(\mathcal{P}_{WE})}$), da seguinte forma:

$$p_{i(\mathcal{P}_{WE})} = \frac{\mathcal{P}_{WE(i)}}{\mathcal{P}_{WE}}, \forall i = 1, \dots, n-1, \quad (3.5)$$

de modo que

$$\sum_{i=1}^{n-1} p_{i(\mathcal{P}_{WE})} = \frac{\sum_{i=1}^{n-1} \mathcal{P}_{WE(i)}}{\mathcal{P}_{WE}} = \frac{\sum_{i=1}^{n-1} n_{WE(i)}}{n_{WE}} \quad (3.6)$$

O fator de ponderação $p_{i(\mathcal{P}_{WE})}$ está dentro do intervalo $[0,1]$.

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

Os valores críticos desejados devem ser encontrados antes de ser realizado o procedimento descrito na Figura (10). O procedimento para estimar o valor crítico de *max-w* é descrito passo a passo a seguir:

1. Especificar a distribuição de probabilidade das estatísticas *w*-test. A f.d.p. atribuída para as estatísticas w_i sob a hipótese nula \mathcal{H}_0 é escrita como:

$$(w_1, w_2, w_3, \dots, w_n)^T \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_w) \quad (3.7)$$

onde $\mathbf{R}_w \in \mathbb{R}^{n \times n}$ é a matriz de correlação entre as estatísticas de teste w_i com os elementos na diagonal principal iguais a 1, e os elementos fora da diagonal são as correlações (ρ_{w_i, w_j}) calculadas por (2.19).

2. O algoritmo Mersenne Twister é empregado para gerar sinteticamente as sequências de números aleatórios uniformemente distribuídos $\mathcal{U}[0, 1]$ e, em seguida, o método Transformação Box-Muller é empregado para converter esses números em uma distribuição normal. Portanto, uma sequência de m vetores aleatórios de w_i são gerados de acordo com a distribuição definida na Etapa 1 (3.7), ou seja:

$$\left[(w_1, w_2, w_3, \dots, w_n)^{T(1)}, (w_1, w_2, w_3, \dots, w_n)^{T(2)}, \dots, (w_1, w_2, w_3, \dots, w_n)^{T(m)} \right] \quad (3.8)$$

3. Para cada sequência em (3.8), calcular o max- w de acordo com (2.17). Assim, teremos:

$$\left(\max_{i \in \{1, \dots, n\}} |w_i|^{(1)}, \max_{i \in \{1, \dots, n\}} |w_i|^{(2)}, \dots, \max_{i \in \{1, \dots, n\}} |w_i|^{(m)} \right) \quad (3.9)$$

4. Classificar em ordem crescente (do menor para o maior) as sequências de max- w em (3.9). Esses valores classificados, denotados por \tilde{w} , fornecem uma representação discreta da função de densidade cumulativa de max- w :

$$\tilde{w}^{(1)} < \tilde{w}^{(2)}, \tilde{w}^{(3)}, \dots, < \tilde{w}^{(m)} \quad (3.10)$$

5. Determinar o valor crítico \hat{k} em função do nível de significância global α' . Esse α' tem o mesmo significado do α' em (2.15). A diferença é que agora a dependência entre as hipóteses é capturada pela matriz de correlação entre os resíduos \mathcal{R}_w :

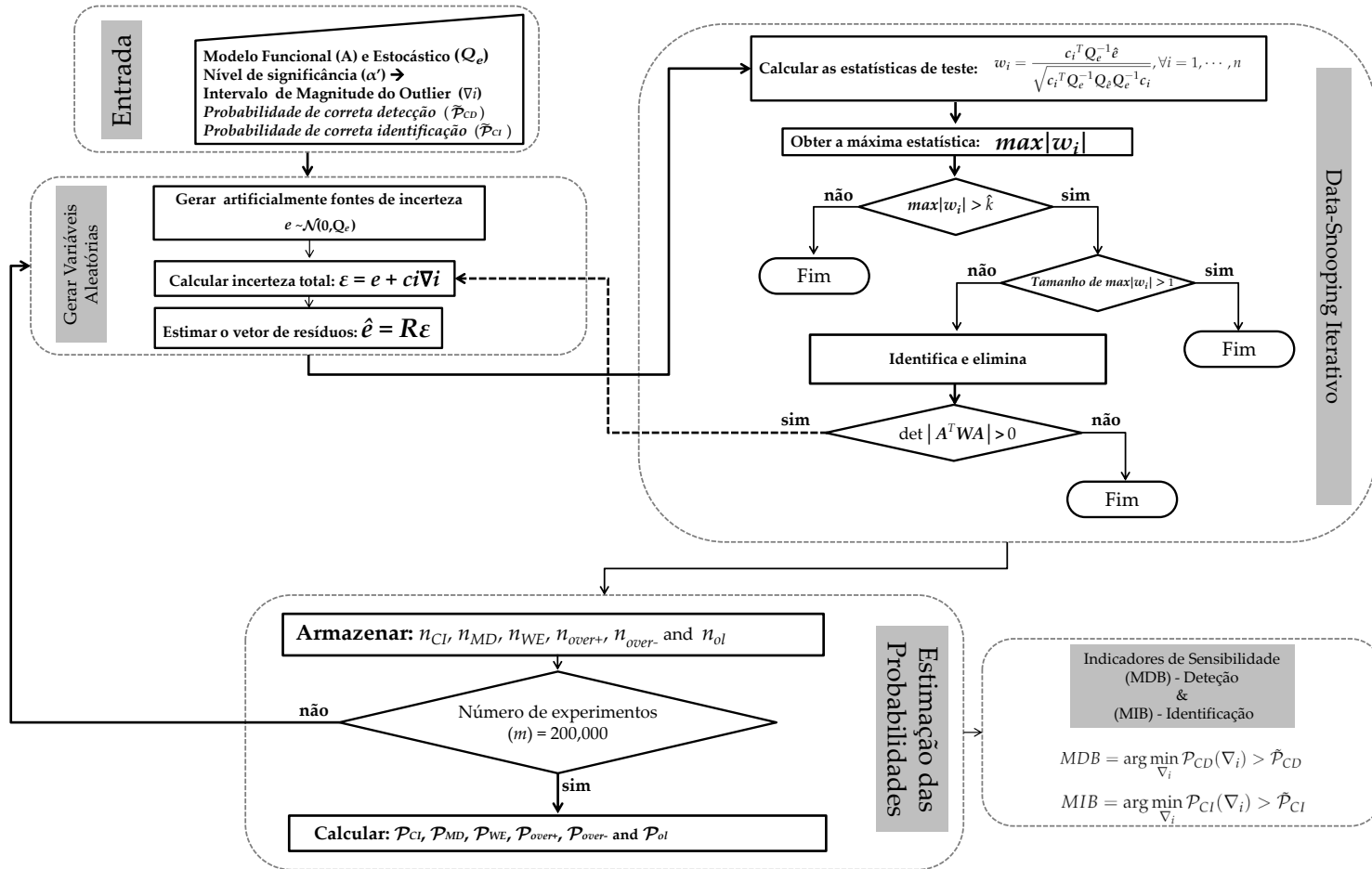
$$\hat{k} = \tilde{w}_{\lceil (1-\alpha') \times m \rceil} \quad (3.11)$$

em que $\lceil \cdot \rceil$ indica arredondamento para o próximo número inteiro e indica a posição dos elementos selecionados na ordem crescente de \tilde{w} . Essa posição corresponde a um valor crítico para um dado α' estipulado pelo usuário. Esse procedimento pode ser feito para uma sequência de α' em paralelo.

3.2 Procedimento para estimação dos níveis de probabilidade do Data-Snooping Iterativo

Após encontrar o valor crítico \hat{k} , a segunda parte do procedimento é executada para estimar os níveis de probabilidade do DSI, conforme ilustrado na Figura (10).

Figura 10 – Fluxograma para estimação das probabilidades do DSI e determinação dos indicadores de sensibilidade (MDB e MIB).



Primeiro, o vetor de incertezas é sinteticamente gerado com base na distribuição normal multivariada. O modelo estocástico para efeitos aleatórios é dado em função da matriz covariância das observações \mathbf{Q}_e . Mersenne Twister e Box-Muller também são utilizados nesse processo.

O intervalo de magnitude do outlier é definido pelo usuário, embora o valor do MDB_0 para uma única hipótese alternativa possa ser utilizado como limite inferior, pois o poder de detecção para uma única hipótese alternativa é menor ou igual ao poder de identificação para múltiplas hipóteses alternativas. As magnitudes são obtidas em função do desvio-padrão das observações, por exemplo $|3\sigma|$ to $|6\sigma|$, em que σ é o desvio-padrão da observação selecionada.

Os outliers podem ser positivos ou negativos. O algoritmo proposto seleciona aleatoriamente o sinal do outlier, de modo que se o valor do sorteio for 1 recebe sinal positivo e se for 0 recebe sinal negativo. O sorteio, portanto possui distribuição uniforme discreta, o que de certa forma, garante que o outlier seja gerado sem tendência de sinal.

Para um único experimento Monte Carlo, teremos um vetor de incertezas $\boldsymbol{\varepsilon}$ composto por dois vetores: um vetor associado aos efeitos aleatórios $\mathbf{e}_i \sim \mathcal{N}(0, \mathbf{Q}_e)$ e outro componente determinístico associado com o outlier $\mathbf{c}_i \nabla_i$.

Em seguida, calculamos o vetor dos resíduos $\hat{\boldsymbol{\varepsilon}}$ conforme a Equação (2.44). Com a disponibilidade do vetor $\hat{\boldsymbol{\varepsilon}}$ estimado, podemos prosseguir com o DSI. O DSI é realizado para m experimentos Monte Carlo, ou seja, m vetores de \mathbf{e}_i , $\forall i = 1, \dots, n$ contaminados por um efeito associado ao outlier $\mathbf{c}_i \nabla_i$.

Com base nos m experimentos Monte Carlo, as probabilidades \mathcal{P}_{CI} , $\mathcal{P}_{CD} = 1 - \mathcal{P}_{MD}$, \mathcal{P}_{WE} , \mathcal{P}_{over+} , \mathcal{P}_{over-} e \mathcal{P}_{ol} podem ser estimadas, respectivamente por:

$$\mathcal{P}_{CI} \approx \frac{n_{CI}}{m} \quad (3.12)$$

$$\mathcal{P}_{MD} \approx \frac{n_{MD}}{m} \quad (3.13)$$

$$\mathcal{P}_{WE} \approx \frac{n_{WE}}{m} \quad (3.14)$$

$$\mathcal{P}_{over+} \approx \frac{n_{over+}}{m} \quad (3.15)$$

$$\mathcal{P}_{over-} \approx \frac{n_{over-}}{m} \quad (3.16)$$

$$\mathcal{P}_{ol} \approx \frac{n_{ol}}{m} \quad (3.17)$$

em que:

- n_{CI} : número de vezes que um outlier é identificado e removido corretamente;
- n_{MD} : número de vezes que um outlier não é detectado;
- n_{WE} : 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 erroneamente identificadas e removidas;
- n_{over-} : número de vezes que duas ou mais observações são identificadas e removidas dos dados observados, enquanto o outlier permanece nos dados observados;
- n_{ol} : número de ocorrências da sobreposição estatística.

Com base nos níveis de probabilidade de correta detecção correta \mathcal{P}_{CD} e correta identificação \mathcal{P}_{CI} , dois índices de sensibilidade do DSI podem ser extraídos: Minimal Identifiable Bias (MIB) e Minimal Detectable Bias (MDB). O MIB expressa a capacidade do DSI em identificar corretamente um outlier, enquanto o MDB está associado com a correta detecção. Esses indicadores podem ser obtidos da seguinte forma:

$$MDB = \arg \min_{\nabla_i} \mathcal{P}_{CD}(\nabla_i) > \tilde{\mathcal{P}}_{CD}, \forall i = 1, \dots, n \quad (3.18)$$

$$MIB = \arg \min_{\nabla_i} \mathcal{P}_{CI}(\nabla_i) > \tilde{\mathcal{P}}_{CI}, \forall i = 1, \dots, n \quad (3.19)$$

A função em (3.18) fornece a magnitude na qual um outlier (∇_i) pode ser detectado, com uma determinada taxa de correta detecção $\tilde{\mathcal{P}}_{CD}$. Por outro lado, a função (3.19) fornece a magnitude na qual um outlier pode ser identificado a partir de uma dada taxa de correta identificação $\tilde{\mathcal{P}}_{CI}$.

Como $\mathcal{P}_{CI} \leq \mathcal{P}_{CD}$, o MIB será sempre maior ou igual ao MDB, isto é, $MIB \geq MDB$. Para o caso especial de se ter uma única hipótese alternativa, não existem diferenças entre o MDB e MIB.

A relação entre o MDB e MIB depende apenas de seus parâmetros de não-centralidade, ou seja:

$$\lambda_{q=1}^{(MDB)} = \frac{MDB_{(i)}^2}{\sigma_{\nabla_i}^2} \quad (3.20)$$

$$\lambda_{q=1}^{(MIB)} = \frac{MIB_{(i)}^2}{\sigma_{\nabla_i}^2} \quad (3.21)$$

$$\frac{MIB_{(i)}}{MDB_{(i)}} = \sqrt{\frac{\lambda_{q=1}^{(MIB)}}{\lambda_{q=1}^{(MDB)}}} \quad (3.22)$$

Importante destacar que no caso do DSI, a correta identificação (\mathcal{P}_{CI}) não depende apenas das probabilidades de falha de detecção (\mathcal{P}_{MD}) e da falsa exclusão (\mathcal{P}_{WE}), mas também das taxas de sobre-identificações (\mathcal{P}_{over+} e \mathcal{P}_{over-}) e as ocorrências de sobreposição estatística (\mathcal{P}_{ol}).

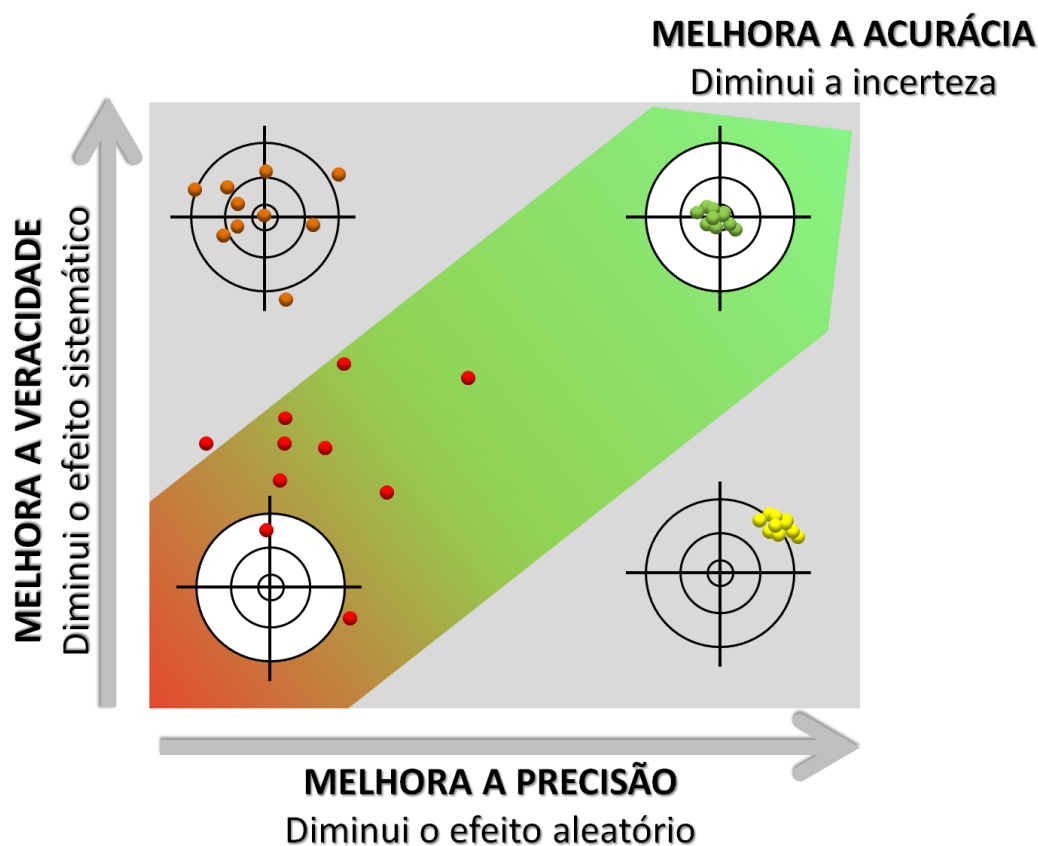
O cálculo para obter o MDB no caso de uma única hipótese alternativa (MDB_0) foi descrito no Capítulo 1. Nota-se pela Equação (2.11) que o MDB_0 possui uma expressão analiticamente fechada (*analytical closed-form expression*), ou seja, pode ser expresso analiticamente em termos de um número delimitado de certas funções bem conhecidas. Entretanto, o cálculo do MIB e MDB para o DSI deve ser feito por meio do Método de Monte Carlo: primeiro porque a região de aceitação, bem como a região crítica, para o caso de múltiplas hipóteses alternativas são analiticamente intratáveis; e segundo porque o DSI envolve múltiplas rodadas consecutivas de estimação, teste e adaptação.

4 Experimentos e Resultados

Neste capítulo, nós apresentamos os resultados na forma de artigos. Eles foram divididos em cinco tópicos de acordo com os objetivos específicos levantados no Capítulo 1, Seção 1.2. O primeiro tópico descreve os principais achados do artigo intitulado “Meio século do conceito de confiabilidade de Baarda: uma revisão, novas perspectivas e aplicações”, da autoria de [Rofatto et al. \(2020a\)](#). Os três tópicos seguintes se originaram do artigo intitulado “Um Método Diagnóstico de Outlier baseado em Monte Carlo para Análise de Sensibilidade”, da autoria de [Rofatto et al. \(2020b\)](#), enquanto que o quinto e último tópico deste capítulo, traz o artigo intitulado “Avaliação dos Efeitos das Injunções Absolutas e Relativas sobre a Sensibilidade do Data-Snooping”, submetido para a revista científica PLOS ONE, de autoria de Vinicius Francisco Rofatto, Marcelo Tomio Martsuoka, Ivandro Klein, Maurício Roberto Veronez e Luiz Gonzaga da Silveira Junior. Os tópicos são descritos a seguir:

- **4.1: Avaliação da acurácia do Método de Monte Carlo.** Até o presente momento, a escolha do número de simulações de Monte Carlo, ou também chamado de número de experimentos Monte Carlo, tem sido avaliada somente em função da precisão. Assim, levanta-se uma questão: como podemos encontrar um número ótimo de experimentos Monte Carlo em termos de acurácia?
- **4.2: Avaliação do valor crítico obtido por meio do Método de Monte Carlo.** O objetivo foi verificar sob quais circunstâncias a dependência entre os resíduos estimados afeta os valores críticos associados à estatística de teste do Data-Snooping.
- **4.3: Análise do efeito do valor crítico controlado sobre os níveis de probabilidade do Data-Snooping.** Aqui, o objetivo foi testar se o aumento do nível de significância conjunto, ou redução do valor crítico, melhora a identificabilidade do outlier.
- **4.4: Estudo da relação entre as medidas analíticas de controlabilidade e os níveis de probabilidade do Data-Snooping.** O objetivo foi estudar a associação entre as medidas analíticas de controlabilidade das observações e os níveis de probabilidade do Data-Snooping, na presença de um outlier dos dados observados.
- **4.5: Avaliação dos efeitos das injunções absolutas e relativas sobre a sensibilidade do Data-Snooping.** Por último, o objetivo foi avaliar o efeito do tipo e quantidade de injunção sobre os indicadores de sensibilidade do Data-Snooping.

Figura 11 – Relação entre Precisão, Veracidade e Acurácia.



Fonte: Produzido pelos autores.

4.1 Avaliação da acurácia do Método de Monte Carlo

Quanto maior o número de experimentos Monte Carlo (m), mais confiável se torna a descrição estatística da variável de interesse. A convergência teórica é da ordem de $m^{-1/2}$. Esta ordem não depende da dimensão dos dados observados (TANIZAKI, 2004). No entanto, até o presente, a escolha do número de experimentos Monte Carlo tem sido avaliada somente em função da precisão, ou seja considerando apenas os efeitos aleatórios. Assim, levanta-se uma questão: como podemos encontrar um número ótimo de experimentos Monte Carlo em termos de acurácia? Acurácia é resultado da veracidade e precisão.

A Figura (11) é fornecida com o objetivo de mostrar a relação entre os termos precisão, veracidade e acurácia. Esses devem ser utilizados sempre que o objetivo for comparar dois ou mais resultados [“mais preciso que”, “mais provável que”, “mais acurado que”]. Se a intenção é expressar o efeito sistemático (tendência), então uma medida adequada deve ser empregada, por exemplo “diferenças absolutas”, “diferenças relativas”, “média da diferença”. Quando um efeito sistemático é reduzido de modo que se torne insignificante para a exatidão requerida pelo problema, apenas efeitos aleatórios estarão presentes. Nesse caso, o investigador pode usar o desvio-padrão (ou variância) para calcular as incertezas de seus resultados, por exemplo JCGM:200:2012 (2012).

Nota-se pela Figura (11) que a acurácia exige o conhecimento prévio da grandeza de interesse. Aqui, usamos eventos com probabilidades conhecidas para avaliar a acurácia do Método de Monte Carlo. Desta forma, pode-se definir o número de experimentos Monte Carlo em função da exatidão requerida. As análises mostram que, dentre os números de experimentos testados, $m = 200,000$ forneceu suficiente precisão numérica, com erro relativo menor que 0.1%. Para mais detalhes veja o tópico **Accuracy of Monte Carlo integration for quality control purposes** do presente artigo.

Com o objetivo de reconhecer o estado da arte sobre as versões da teoria da confiabilidade em geodésia, o artigo também traz uma visão sistêmica dos trabalhos produzidos desde a época de Baarda (1968) até os dias de hoje.

A partir da análise de 60 obras publicadas nos principais periódicos de geodésia, traçamos novas possibilidades e perspectivas de pesquisas, com particular ênfase para o Método de Monte Carlo. Como em qualquer área que depende da tecnologia, notamos que a evolução dos computadores foi a principal força motriz dos avanços da teoria da confiabilidade. O Quadro 1 fornece um panorama geral da evolução da Teoria da Confiabilidade para a geodésia.

No que diz respeito ao procedimento computacional para estimar os níveis de probabilidade do DSI, o método descrito no capítulo anterior está em sua terceira versão (versão mais recente). Como visto, fornecemos uma rotina baseada em Monte Carlo na qual o valor crítico da estatística associada ao DSI é obtido em função do modelo funcional e estocástico do problema. Conseqüentemente, as dependências entre os resíduos dos mínimos quadrados são capturadas pela simulação de Monte Carlo. Neste caso, o usuário tem efetivo controle sobre o Erro Tipo I. Além disso, foi proposto uma nova classe de probabilidade que não havia nas versões anteriores. Em sua primeira versão, os outliers eram sorteados uniformemente e, portanto, o que se tinha eram médias das probabilidades. Além disso, os níveis de probabilidade do DSI foram limitados à quatro classes (\mathcal{P}_{CI} , \mathcal{P}_{CD} , \mathcal{P}_{over+} e \mathcal{P}_{over-}). Detalhes da primeira versão pode ser vista em Rofatto, Matsuoka e Klein (2017). Em sua segunda versão, os níveis de probabilidade associados com o DSI foram calculados para cada observação individual. A segunda versão é descrita no tópico **Probabilities of IDS based on Monte Carlo integration**, do artigo fornecido nesta seção.

Quadro 1 – Sumário da evolução da teoria da confiabilidade.

Autores	Título do Trabalho	Comentários
(BAARDA, 1967)	Statistical concepts in geodesy	Aplicações elementares de métodos estatísticos para problemas de ajustamento em geodésia.
(BAARDA, 1968)	A testing procedure for use in geodetic networks	Introduz o conceito de confiabilidade de redes geodésicas
(MIERLO, 1975)	Statistical analysis of geodetic networks designed for the detection of crustal movements	Primeiras ideias sobre separabilidade de hipóteses alternativas.
(FÖRSTNER, 1983)	Reliability and discernability of extended Gauss-Markov models	Desenvolveu a teoria de separabilidade, considerando duas hipóteses alternativas..
(SCHAFFRIN, 1997)	Reliability Measures for Correlated Observations	Apresenta medidas de confiabilidade para observações correlacionadas.
(AYDIN; DEMIREL, 2004)	Computation of Baarda's lower bound of the non-centrality parameter	Fornecem dois algoritmos recursivos para o cálculo do parâmetro de não-centralidade.
(KNIGHT; WANG; RIZOS, 2010)	Generalised measures of reliability for multiple outliers	Estendem a teoria de confiabilidade para o caso de múltiplos (simultâneos) outliers
(LEHMANN, 2012)	Improved critical values for extreme normalized and studentized residuals in Gauss-Markov models	Otimiza os valores críticos do w -test e τ -test com base no Método Monte Carlo.
(YANG et al., 2013)	Outlier separability analysis with a multiple alternative hypotheses test	Estendem a teoria de confiabilidade e apresentam uma proposta para estimar o poder do teste mínimo do <i>Data-Snooping</i> , bem como valores limites para o nível de confiança
(PRÓSZYŃSKI, 2015)	Revisiting Baarda's concept of minimal detectable bias with regard to outlier identifiability	Propõe índices de identificabilidade, a saber: 1. índices relacionados à identificação de uma observação contaminada dentro de um conjunto de observações suspeitas de conter um outlier. 2. índices relacionados à identificação de uma observação contaminada dentro de um conjunto de observações.
(TEUNISSEN, 2018)	Distributional theory for the DIA method	Apresenta uma teoria que unifica o procedimento de estimativa e teste do método DIA (Detecção, Identificação e Adaptação).
(IMPARATO; TEUNISSEN; TIBERIUS, 2019)	Minimal Detectable and Identifiable Biases for quality control	Analisa a diferença entre o indicador de menor outlier identificável (MIB) e o menor outlier detectável (MDB). Mostram que, sob múltiplas hipóteses alternativas, $MIB \geq MDB$.
(ZAMINPARDAZ; TEUNISSEN, 2019a)	DIA-datasnooping and identifiability	Analisa o <i>Data-Snooping</i> no contexto do DIA. Investiga as circunstâncias em que duas ou mais hipóteses não são separáveis na etapa de identificação.

A half-century of Baarda's concept of reliability: A review, new perspectives, and applications

Over the 50 years of its existence, Baarda's concept of reliability has been used as a standard practice for the quality control in geodesy and surveying. In this study, we analysed the pioneering work of Baarda (Publ Geod New Ser 2(4) 1967; Publ Geod New Ser 2(5) 1968) and recent studies on the subject. We highlighted that the advent of personal computers with powerful processors has rendered simulation as an attractive and cost-effective approach for quality control purposes. We also provided an overview of the latest advances in the reliability theory for geodesy, with particular emphasis on Monte Carlo simulation.

Keywords: Reliability, Outlier, Hypothesis Testing, Quality Control, Minimal Detectable Bias, Minimal Identifiable Bias, Data snooping

Inspiration and Motivation

In many geodetic analysis tasks, large sets of observations are being recorded or sampled. It is nearly impossible that such datasets are free from outliers. Thus, it is important to identify outlying observations that may lead to model misspecification, biased parameter estimation, and incorrect results. Among the many definitions found in the literature on what an outlier is, we follow the statement given by Lehmann (2013): 'an outlier is an observation that is so probably caused by a gross error that it is better not used or not used as it is'.

Over the 50 years of its existence, Baarda's (1967, 1968) groundbreaking work on outlier testing has been used as a standard practice for the quality control in geodesy and surveying. As part of his contribution, he proposed a procedure based on hypothesis testing for the detection of a single outlier in linear(ised) models, which he called data snooping. Most of conventional geodetic studies have a chapter on Baarda's data snooping procedure, e.g. Koch (1999) and Teunissen (2006). Furthermore, this

procedure has become very popular and is routinely used in adjustment computations (Ghilani 2010). Although data snooping was introduced as a testing procedure for use in geodetic networks, it is a generally applicable method (Lehmann 2012).

Baarda's data snooping consists of screening each individual observation for a possible outlier (Kok 1984, Teunissen 2006). Baarda's w -test statistic for his data snooping is given by a normalised least-squares residual. This test, which is based on a linear mean-shift model, can also be derived as a particular case of the generalised likelihood ratio test (Kok 1984, Koch 1999, Teunissen 2006). In principle, Baarda's w -test only makes a decision between the null \mathcal{H}_0 and a single alternative hypothesis \mathcal{H}_i . In that case, rejection of \mathcal{H}_0 automatically implies acceptance of \mathcal{H}_i , and vice versa (Teunissen 2017, Imperato *et al.* 2018). Therefore, it concerns on outlier detection and not on outlier identification (Teunissen 2017).

Based on the probability of rejecting a true null hypothesis (type I error - 'false alarm', denoted by α) and the probability of rejecting a true alternative hypothesis (type II error - 'missed detection', denoted by β), Baarda (1968) also derived the concept of a marginally detectable outlier (Kok 1984), which is commonly called as the minimal detectable bias (MDB)—the term given by Teunissen (1989). The MDB of an alternative hypothesis is the smallest outlier that can lead to rejection of a null hypothesis for a given α and β . Thus, for each alternative hypothesis \mathcal{H}_i , the corresponding MDB can be easily computed. The set of MDBs describes the internal reliability, whereas their propagation into the parameters is said to describe the external reliability (Baarda 1968, Vaniček *et al.* 2001, Teunissen 2006).

Different from the case of the single alternative hypothesis above, Förstner (1983) developed the separability analysis between two alternative hypotheses, e.g. \mathcal{H}_i and \mathcal{H}_j . In this case, rejection of \mathcal{H}_0 does not necessarily imply the correct identification

of a particular alternative hypothesis. Consequently, there is also a probability of correctly rejecting the null hypothesis but wrongly accepting an alternative hypothesis. This type of wrong decision is called type III error (Hawkins 1980) or ‘wrong alarm’ (indicated by ‘ κ ’ in this paper). In this respect, we discuss the identification of an outlier. For identification, the correlation coefficient (here, denoted by ρ) between Baarda's w -test should be considered during the data snooping procedure (Förstner 1983, Wang and Knight 2012, Yang *et al.* 2013, Prószyński 2015).

Investigations on outlier identification involve separability analyses between alternative hypotheses. Generally, a separability analysis reveals the degree to which the alternative hypotheses can be distinguished. Van Mierlo's (1975, 1979) studies were the first to show the possibility of describing the separability of alternative hypotheses on the so-called wrong alarm in deformation monitoring. The current outlier separability theory is credited to Yang *et al.* (2013). They have extended the separability analysis from two alternative hypotheses (Förstner 1983, Li 1986) to multiple alternative hypotheses. Thus, for a system with n measurements, the null hypothesis \mathcal{H}_0 should be tested against all n alternative hypotheses.

Recently, Prószyński (2015) provided supplementation of the MDB by an outlier identifiability index to each observation and misidentifiability index as the maximum probability of wrong alarm rate. Teunissen (2017) introduced the concept of minimal identifiable bias (MIB) as the smallest outlier of an alternative hypothesis that can lead to its identification for a given probability. Imperato *et al.* (2018) showed that, under multiple testing, the MIBs should be larger than their MDBs in order to guarantee a successful outlier identification. There is no difference between MDB and MIB when only a single alternative hypothesis is involved (Teunissen 2017, Imperato *et al.* 2018). This means that an outlier with a magnitude of MDB will hardly be identified. In n

alternative hypotheses, therefore, the MDB should increase with the correlation coefficient (Yang *et al.* 2013).

The articles quoted above, i.e. Yang *et al.* (2013), Prószczyński (2015), and Imparato *et al.* (2018), have already pointed out that the density functions and the critical region under the multiple alternative hypotheses are extremely complex and impossible to obtain in a closed form or even by numerical integration. They used an empirical method based on numerical simulation of random observation errors to analyse their analytical expression. The situation becomes a little more complicated when Baarda's data snooping is applied in its iterative form. This procedure is called iterative data snooping (IDS) (Teunissen 2006, p. 135). In contrast to the well-defined theories (Baarda 1968, Förstner 1983, Wang and Knight 2012, Yang *et al.* 2013, Prószczyński 2015, Imparato *et al.* 2018), we believe that the IDS procedure is a heuristic method, and therefore, there is no complete and rigorous reliability theory on it. Heuristics are understood as techniques designed for finding approximate solutions when classic methods fail to find any exact solution. In such case, an analytical model with a tractable solution for its probabilities is unknown, and therefore, one needs to resort to computer simulations such as Monte Carlo methods.

When the underlying system is extremely complex, such as in IDS, a numerical simulation is generally preferable for many reasons. For instance (Altiok and Melamed 2007)

- (1) the time spent by researchers (engineer, designer, surveyor) to derive a solution is excessive;
- (2) the researcher judges that an attempt at an analytical solution is a poor approach owing to the apparent mathematical difficulties;

- (3) the researcher cannot even formulate an analytical expression with sufficient power to capture the system's behavioural aspects of interest.

One can argue on the disadvantage of high computational complexity involved in a simulation approach. In contrast to Baarda's era, we currently have fast computers and large data storage systems at our disposal. Therefore, such obstacles are no longer existing because computing power is not a bottleneck at present (Lehmann 2015). In this context, the Monte Carlo simulation has been extensively applied in geodesy (Lehmann 1994, Hekimoglu and Koch 1999, Koch 2007, Lehmann and Scheffler 2011, Aydin 2012, Lehmann 2012, Yang *et al.* 2013, Koch 2016, Klein *et al.* 2015, Prószyński 2015, Marx 2015, Lehmann and Voß-Böhme 2017, Klein *et al.* 2017, Rofatto *et al.* 2017, Imperato *et al.* 2018). Monte Carlo simulation models incorporate randomness by sampling random values from specified distributions. The basic idea is to approximate probability distributions by frequency distributions of computer random experiments performed using pseudo-random number generators (PRNGs) (Koch 2007, Lehmann 2012).

Despite the countless contributions made over the last 50 years, there is continuing research on the topic, which paves the way for further investigations. In addition to the Monte Carlo simulation, new tools have become available, such as genetic algorithms (Alma *et al.* 2011), swarm intelligence (Ye and Chen 2008), simulated annealing (Weisberg and Atkinson 1991), fuzzy approaches (Yousri *et al.* 2007), and metaheuristic algorithms (Koch *et al.* 2017), to handle datasets contaminated by outliers. Their potential for outlier treatment is not yet fully exploited in geodesy.

The purpose of this study is to provide an overview of the latest advances in the reliability theory for geodesy. Here, we analysed the pioneering work of Baarda (1968) and the recent studies on the subject (Lehmann 2012, Yang *et al.* 2013, Prószyński

2015, Teunissen 2017, Imparato *et al.* 2018). Then, we pointed out new possibilities and perspectives on the issue, with a particular emphasis on the Monte Carlo simulation.

The paper is organised as follows. First, we provide the elements of Baarda's original data snooping testing procedure and its restrictions when there are multiple alternative hypotheses. Second, we demonstrate by an illustrative example how IDS works, and we point out that its probabilities should be computed by a simulation technique. Third, we present the Monte Carlo simulation method as an alternative for quality control. With regard to the Monte Carlo technique, two problems are presented and we try to answer them. First, how can we find the optimal number of Monte Carlo experiments for quality control purpose? Second, how do we obtain the probability levels of IDS? Finally, the concluding remarks are summarised at the end of this paper.

Baarda's data snooping for outlier detection

Null hypothesis versus alternative hypothesis: a binary case

The null hypothesis, which is also called the working hypothesis, corresponds to a supposedly valid model describing the physical reality of the observations without the presence of an outlier. When it is assumed to be ‘true’, this model is used to estimate the unknown parameters, typically in a least-squares approach. Thus, the null hypothesis of the standard Gauss–Markov model in linear or linearised form is given by (Koch 1999)

$$\mathcal{H}_0: E\{\mathbf{y}\} = \mathbf{A}\mathbf{x}, D\{\mathbf{y}\} = \Sigma_{\mathbf{y}\mathbf{y}} \quad (1)$$

where $E\{\cdot\}$ is the expectation operator, $\mathbf{y} \in \mathbb{R}^n$ the vector of measurements, $\mathbf{A} \in \mathbb{R}^{n \times u}$ the Jacobian matrix (also called design matrix) of full rank u , $\mathbf{x} \in \mathbb{R}^u$ the unknown parameter vector, $D\{\cdot\}$ the dispersion operator, and $\Sigma_{\mathbf{y}\mathbf{y}} \in \mathbb{R}^{n \times n}$ the known positive definite variance matrix of the measurements. In this case, the redundancy r of the

model in Eq. (1) is $r = n - u$.

Instead of \mathcal{H}_0 , Baarda (1968) proposed an alternative hypothesis \mathcal{H}_A , also referred to as model misspecification by Teunissen (2006), as follows:

$$\mathcal{H}_A: E\{\mathbf{y}\} = \mathbf{A}\mathbf{x} + \mathbf{C}\mathbf{v}, D\{\mathbf{y}\} = \Sigma_{yy} \quad (2)$$

with $\mathbf{C} \in \mathbb{R}^{n \times q}$ as the matrix of outlier model and $\mathbf{v} \in \mathbb{R}^q$ the unknown outlier vector.

Under \mathcal{H}_A , we restrict ourselves to $[\mathbf{A} \quad \mathbf{C}] \in \mathbb{R}^{n \times (u+q)}$ having a full column rank, such that $q \leq n - u = r$. The number of unknown outliers defines the alternative model, i.e.

$$\overbrace{1}^{\text{local test}} \leq q \leq \overbrace{n-u}^{\text{overall test}} \quad (3)$$

When $q = n - u = r$, an overall model test is performed. In that extreme case, the redundancy of the alternative hypothesis \mathcal{H}_A is $r = n - (u + q) = n - (u + n - u) \therefore r = 0$, and therefore, there is no need to specify any particular alternative hypothesis. In this case, the overall model test is based on the weighted sum-of-squared residuals $\|\hat{\mathbf{e}}_0\|^2$, i.e.

$$\|\hat{\mathbf{e}}_0\|^2 = \hat{\mathbf{e}}_0^T \Sigma_{yy}^{-1} \hat{\mathbf{e}}_0 \quad (4)$$

with $\hat{\mathbf{e}}_0$ being the least-squares residuals vector under \mathcal{H}_0 . By setting an acceptance region (also known as size of test) $\mathcal{A}_{\alpha_0} \subset \mathbb{R}^n$ (Baarda 1968, Teunissen 2006),

$$\mathcal{A}_{\alpha_0} = \{\mathbf{y} \in \mathbb{R}^n \mid \|\hat{\mathbf{e}}_0\|^2 \leq \chi_{\alpha_0}^2(r, 0)\} \quad (5)$$

Thus, the overall test is given as

$$\text{Accept } \mathcal{H}_0 \text{ if } \|\hat{\mathbf{e}}_0\|^2 \leq \chi_{\alpha_0}^2(r, 0), \text{ reject otherwise} \quad (6)$$

The critical value $\chi_{\alpha_0}^2(r, 0)$ is computed from the central chi-squared distribution with

$r = n - u$ degrees of freedom and a chosen type I error, also known as false alarm or level of significance, α_0 (note: the index ‘0’ represents the univariate probability). The performance of that test can be described by its probabilities of committing type I errors and type II errors (also known as missed detection rate, β_0) as follows:

$$\begin{aligned} \text{Type I error: } & \mathbf{P}[\mathbf{y} \notin \mathcal{A}_{\alpha_0} | \mathcal{H}_0 = \text{true}] = \alpha_0 \\ \text{Type II error: } & \mathbf{P}[\mathbf{y} \in \mathcal{A}_{\alpha_0} | \mathcal{H}_0 = \text{false}] = \beta_0 \end{aligned} \quad (7)$$

The probability of type I error α_0 is the probability of rejecting the null hypothesis when it is true, whereas the type II error β_0 is the probability of failing to reject the null hypothesis when it is false. The complement of type II error is the well-known power of the test (also known as correct detection) given by $\gamma_0 = 1 - \beta_0$.

For the overall model test, the weighted sum-of-squared residuals under \mathcal{H}_A follow the non-central chi-squared distribution with $r = n - u$, i.e.

$$\|\hat{\mathbf{e}}_A\|^2 \sim \chi_{\alpha_0}^2(r, \lambda_q), \text{ where } \lambda_q = \mathbf{C}^T \mathbf{\nabla}^T \Sigma_{yy}^{-1} \Sigma_{\hat{\mathbf{e}}_0} \Sigma_{yy}^{-1} \mathbf{C} \mathbf{\nabla} \quad (8)$$

where $\hat{\mathbf{e}}_A$ is the least-squares residuals vector under \mathcal{H}_A , λ_q the non-centrality parameter, and $\Sigma_{\hat{\mathbf{e}}_0}$ the variance matrix of the best linear unbiased estimator of $\hat{\mathbf{e}}_0$ under \mathcal{H}_0 . The non-centrality parameter λ_q describes the discrepancy between the null \mathcal{H}_0 and \mathcal{H}_A . We restrict ourselves to the data snooping by Baarda (1968). For more details about the overall model test, see Teunissen (2017) and Imparato *et al.* (2018).

On the other hand, Baarda (1968) proposed a procedure to detect a single outlier in linear(ised) models, which he called data snooping. The purpose of his data snooping procedure is to screen each individual observation for an outlier. Therefore, in contrast to the overall model test, Baarda's data snooping is based on a local model test. In such case, $q = 1$ and the n alternative hypothesis is given by

$$\mathcal{H}_i: E\{\mathbf{y}\} = \mathbf{A}\mathbf{x} + c_i\nabla_i (\forall i = 1, \dots, n), D\{\mathbf{y}\} = \Sigma_{yy} \quad (9)$$

Now, the matrix \mathbf{C} in Eq. (2) is reduced to a canonical unit vector c_i , which consists exclusively of elements with values of 0 and 1, where 1 means that an i th outlier of magnitude ∇_i affects an i th measurement and 0 otherwise. In that case, the rank $[\mathbf{A} \quad c_i] \in \mathbb{R}^{n \times (u+1)}$ and the vector ∇ in Eq. (2) reduces to a scalar ∇_i in Eq. (9), i.e.

$$c_i = \begin{bmatrix} 0 & 0 & \dots & \overset{i}{\hat{1}} & 0 & \dots & 0 \end{bmatrix}^T.$$

Baarda's w -test statistic for his data snooping is given by a normalised least-squares residual as follows (Baarda 1968):

$$w_i = \frac{c_i \Sigma_{yy}^{-1} \hat{\mathbf{e}}_0}{\sqrt{c_i^T \Sigma_{yy}^{-1} \Sigma_{\hat{\mathbf{e}}_0} \Sigma_{yy}^{-1} c_i}} (\forall i = 1, \dots, n) \quad (10)$$

To verify if there are sufficient evidences to reject or not the null hypothesis, the test for binary case should be performed as

$$\text{Accept } \mathcal{H}_0 \text{ if } |w_i| \leq \sqrt{\chi_{\alpha_0}^2(r=1,0)} = \sqrt{k}, \text{ reject otherwise in favour of } \mathcal{H}_i \quad (11)$$

with the acceptance region given as

$$\mathcal{A}_{\alpha_0}^i = \{\mathbf{y} \in \mathbb{R}^n \mid |w_i| \leq \sqrt{k}\} \quad (12)$$

with k being the critical value from the chi-squared one-side distribution (note: the square root of a chi-squared one-side test is equivalent to that of a normal distribution two-tailed test). For example, for $\alpha_0 = 0.01$, we obtain $\sqrt{k} = 2.576$. In this case, if $|w_i| > 2.576$ for some y_i , one may reject \mathcal{H}_0 because $y_i \notin \mathcal{A}_{\alpha_0}^i$.

Because Baarda's w -test in its essence is based on binary hypothesis testing, in which one decides between the null hypothesis \mathcal{H}_0 and a unique alternative hypothesis

\mathcal{H}_i of Eq. (9), it may lead to type I errors (α_0) and type II errors (β_0). Instead of α_0 and β_0 , there is the confidence level (CL) and power of the test (γ_0), respectively. The first deals with the probability of accepting a true null hypothesis; the second, as already mentioned, with the probability of correctly accepting the alternative hypothesis.

The normalised least-squares residual w_i follows a standard normal distribution with the expectation that $\mu_0 = 0$ if \mathcal{H}_0 holds true (there is no outlier). On the other hand, if the system is contaminated with a single outlier at the i th position of the dataset (i.e. under \mathcal{H}_i), there is an outlier that causes the expectation of w_i to become $\lambda_{q=1}$ (i.e. $\mu_i = \lambda_{q=1}$), which is the non-centrality parameter for $q = 1$. In this case, the non-centrality parameter $\lambda_{q=1}$ is given by

$$\lambda_{q=1} = \mathbf{c}_i^T \Sigma_{\mathbf{y}\mathbf{y}}^{-1} \Sigma_{\hat{\mathbf{e}}_0} \Sigma_{\mathbf{y}\mathbf{y}}^{-1} \mathbf{c}_i \nabla_i^2 \quad (13)$$

Fig. 2 shows the relationship among the parameters α_0 , β_0 , CL, γ_0 , $\lambda_{q=1}$, and $\sqrt{k} = \sqrt{\chi_{\alpha_0}^2(r=1,0)}$. For a given value of $\lambda_{q=1}$, the smaller the significance level α_0 , the smaller the power of the test γ_0 . On the other hand, the larger the non-centrality parameter $\lambda_{q=1}$, the larger the power γ_0 for a given α_0 .

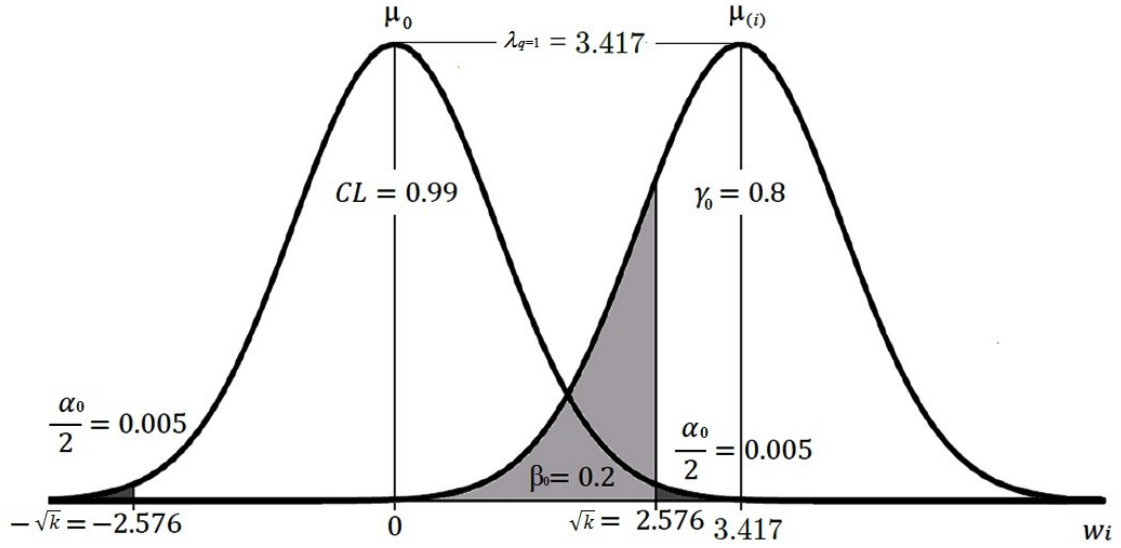


Figure 1. Probability levels fixed at $CL = 0.99$ ($\alpha_0 = 0.01$) and $\gamma_0 = 0.8$ ($\beta_0 = 0.2$), which lead to a non-centrality parameter $\lambda_{q=1} = 3.417$ and $\sqrt{k} = 2.576$.

The non-centrality parameter $\lambda_{q=1}$ in Eq. (13) describes the discrepancy between the null \mathcal{H}_0 and the single alternative hypothesis \mathcal{H}_i in Eq. (9). In other words, the non-centrality parameter $\lambda_{q=1}$ represents the expected mean shift of a specific w -test when there is an outlier. In such case, the term $\mathbf{c}^T \Sigma_{yy}^{-1} \Sigma_{\hat{e}_0} \Sigma_{yy}^{-1} \mathbf{c}_i$ becomes a scalar and the solution of the quadratic Eq. (13) is given by (Teunissen 2006)

$$|\nabla_i| = \text{MDB}(i) = \sqrt{\frac{\lambda_{q=1}(\alpha_0, \gamma_0, r = q = 1)}{\mathbf{c}_i^T \Sigma_{yy}^{-1} \Sigma_{\hat{e}_0} \Sigma_{yy}^{-1} \mathbf{c}_i}} \quad (14)$$

where $|\nabla_i|$ is $\text{MDB}(i)$, which is computed for each of the n alternative hypotheses according to Eq. (9).

Because

$$\sigma_{\hat{v}_i}^2 = (\mathbf{c}_i^T \Sigma_{yy}^{-1} \Sigma_{\hat{e}_0} \Sigma_{yy}^{-1} \mathbf{c}_i)^{-2} \quad (15)$$

with $\sigma_{\hat{v}_i}^2$ being the variance of estimated outlier \hat{v}_i , $\text{MDB}(i)$ can also be computed by combining Eqs. (14) and (15) as

$$\text{MDB}(i) = \sigma_{\hat{v}_i} \sqrt{\lambda_{q=1}(\alpha_0, \gamma_0, r = q = 1)}, (\forall i = 1, \dots, n) \quad (16)$$

where $\sigma_{\hat{v}_i}$ is the standard deviation of the estimated outlier \hat{v}_i .

MDB is the smallest outlier that can be detected with the least probability $\gamma_0 = 1 - \beta_0$ (Baarda 1968, Kok 1982, Teunissen 1989). Note that the non-centrality parameter $\lambda_{q=1}$ in Eq. (13) requires knowledge of the outlier size, which in practice is unknown. On the other hand, $\lambda_{q=1}$ can be computed as a function of α_0 , β_0 , and $r = q = 1$. Baarda (1968) provided monograms for those interested in obtaining values as a function of α_0 , β_0 , and r . Alternatively, we use the recursive algorithm based on the work by Aydin and Demirel (2005), namely Newton algorithm, in order to obtain the non-centrality parameter $\lambda_{q=1}$. The left side graph of Fig. 2 shows, for the case $q = 1$, that MDB becomes larger for smaller values of α_0 and β_0 (or larger power of the test γ_0). It should also be noted that the more precise the estimated \hat{v}_i (smaller $\sigma_{\hat{v}_i}$), the smaller the MDB (see right graph of Fig. 2).

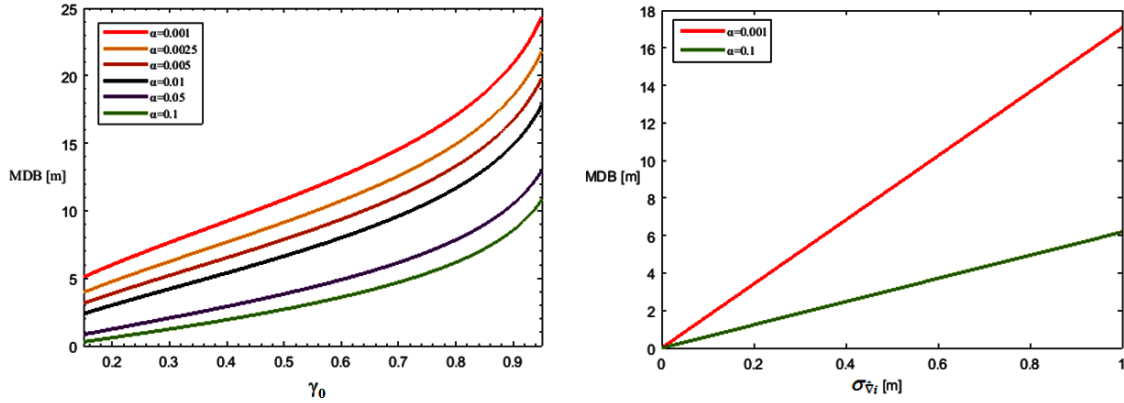


Figure 2. Left: MDB computed as a function of power of the test γ_0 for different values of type I error α_0 (the estimation of the outlier precision was fixed as $\sigma_{\hat{v}_i} = 1$ m). Right: MDB as a function of $\sigma_{\hat{v}_i}$ for $\alpha_0 = 0.001$ and $\alpha_0 = 0.1$ (the power of the test was set as $\gamma_0 = 0.8$).

MDB is the lower bound of the outlier that can be successfully detected. It was further investigated for a single outlier in a singular Gauss–Markov model (Wang and

Chen 1999). There are also studies covering either independent or correlated measurements (Pelzer 1983, Prószczyński 1994, Prószczyński 2010, Wang and Chen 1994, Schaffrin 1997, Teunissen 1998).

For the case of $1 < q < n - u$, one assumes that more than one outlier is present in the dataset. In other words, it is possible to set up for the case of multiple outliers (Kok 1984, Ding and Coleman 1996, Teunissen 2006, Gui *et al.* 2011, Knight *et al.* 2010, Baselga 2011, Koch 2016, Klein *et al.* 2017). The multiple outlier case is covered at full length by Teunissen (2006) [For more details about alternative models, refer to Lehmann (2013) and Lehmann and Lösler (2016)].

Although Baarda's w -test belongs to the class of generalised likelihood ratio tests and has the property of being a uniformly most powerful invariant (UMPI) test when the null hypotheses is tested against a single alternative (Arnold 1981, Teunissen 2006, Kargoll 2007), this test may not necessarily be a UMPI when more than one alternative hypothesis are considered, as is the case of the data snooping procedure (Kargoll 2012, Lehmann and Voß-Böhme 2017, Teunissen 2017).

Baarda's data snooping for outlier identification

In this section, we will briefly review the two alternative hypotheses case (Förstner, 1983) and the case of n alternative hypotheses (Yang *et al.* 2013, Prószczyński 2015, Teunissen 2017, Imperato *et al.* 2018).

Förstner method: Two alternative hypotheses

After 15 years of existence, Baarda's theory was extended by Förstner (1983), who considered two alternative hypotheses instead of a single one. Under two alternative hypotheses, the non-centrality parameter is not only related to the sizes of α_0 and β_0 as shown until the present, but it is also dependent on the correlation coefficient (ρ)

between Baarda's w -test statistics, i.e. w_i and w_j , as follows (Förstner 1983, Li 1986):

$$\rho_{w_i, w_j} = \frac{\mathbf{c}_i^T \Sigma_{yy}^{-1} \Sigma_{\hat{e}_0} \Sigma_{yy}^{-1} \mathbf{c}_j}{\sqrt{\mathbf{c}_i^T \Sigma_{yy}^{-1} \Sigma_{\hat{e}_0} \Sigma_{yy}^{-1} \mathbf{c}_i} \sqrt{\mathbf{c}_j^T \Sigma_{yy}^{-1} \Sigma_{\hat{e}_0} \Sigma_{yy}^{-1} \mathbf{c}_j}} \quad (i \neq j) \quad (17)$$

Different from the MDB presented in the previous section, the correlation coefficient ρ_{w_i, w_j} in Eq. (17) does not depend on the non-centrality parameter, which is a function of the probability levels α_0 , β_0 , and r . Actually, it depends only on the design model (matrices \mathbf{A} and \mathbf{W}) and the outlier models \mathbf{c}_i and \mathbf{c}_j . \mathbf{W} is the weight matrix of the observations, taken as $\mathbf{W} = \sigma_0^2 \Sigma_{yy}^{-1}$, where σ_0^2 is the variance factor (Koch 1999, Lehmann 2012).

For the test with two alternative hypotheses, apart from type I (false alarm) and type II errors (missed detection), there is a third type of wrong decision when Baarda's data snooping is performed. Baarda's data snooping can also flag a non-outlying observation while the 'true' outlier remains in the dataset. We are referring to the type III error (Hawkins 1980), also called 'wrong alarm' or 'wrong exclusion' (Van Mierlo 1979, Förstner 1983, Li 1986, Yang *et al.* 2017). The determination of the type III error (here denoted by κ_{ij}) involves a separability analysis between the alternative hypotheses (Van Mierlo 1979, Förstner 1983, Li 1986, Wang and Knight 2012, Yang *et al.* 2013, Prószyński 2015, Yang *et al.* 2017). Therefore, we are now interested in the identification of the correct alternative hypothesis. Suppose that there is an outlier in the i th position of the dataset, the probability of committing a type III error (κ_{ij}) is given by

$$\text{Type III error: } \mathbb{P}[|w_j| > |w_i|, |w_j| > \sqrt{k} \mid (i \neq j) \mid \mathcal{H}_i = \text{true}] = \kappa_{ij} \quad (18)$$

In contrast to the single alternative hypothesis, the acceptance region for the case of two alternative hypotheses, i.e. \mathcal{H}_i and \mathcal{H}_j , is now given by the intersection of two

individual acceptance regions, i.e. $\mathcal{A}_{\alpha_{00}} = \mathcal{A}_{\alpha_0}^i \cap \mathcal{A}_{\alpha_0}^j (i \neq j)$. In this case, the intersection of the two individual acceptance regions has a smaller area compared to each of the individual ones (note: the index ‘00’ denotes the bivariate probability, whereas ‘0’ represents the univariate probability). Thus, the type II error probability becomes smaller, i.e. $\beta_{00} \leq \beta_0$, whereas the type I error becomes larger, i.e. $\alpha_{00} > \alpha_0$. Now, the power of the test is the complement of the sum of type II and III errors, $\gamma_0 = 1 - (\beta_{00} + \kappa_{ij})$. Furthermore, the probabilities of committing type I, II, and III errors are essentially related to the dependencies of the alternative hypotheses. In this case, the larger the correlation coefficient, the larger the type III error and the smaller the type II error (Yang *et al.* 2013).

Based on the work by Yang *et al.* (2013), the left side graph of Fig. 3 shows that α_{00} becomes smaller when ρ increases from 0 to 1, given $\alpha_0 = 0.01$ and $\beta_{00} + \kappa_{ij} = 0.2$. The middle graph of Fig. 3 shows that the larger the correlation coefficient ($\rho > 0.8$), the higher the type III error (κ_{ij}). When the correlation coefficient is very large, β_{00} is effectively zero whereas κ_{ij} is 0.2 (20.0%). Therefore, the larger the correlation coefficient, the larger the value of κ_{ij} and the smaller the β_{00} . The right side graph of Fig. 3 shows that the non-centrality parameter λ_{00} for two alternative hypotheses increases as the correlation increases from 0 to 1.

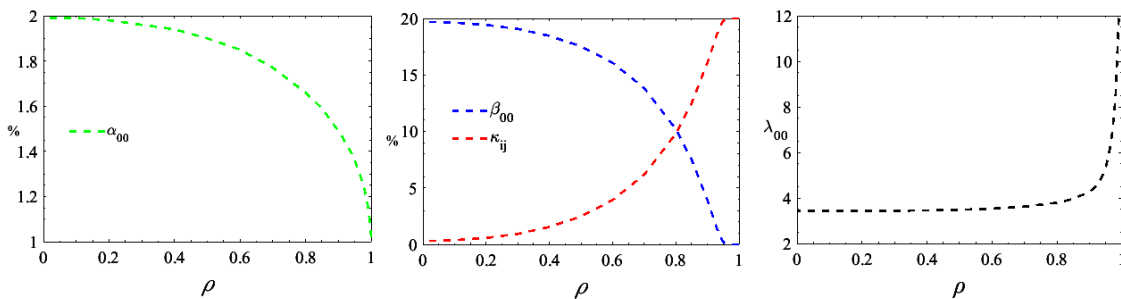


Figure 3. Left: $\uparrow \alpha_{00}$ ($\downarrow \sqrt{k}$) if $\rho \downarrow$; Middle: $\downarrow \beta_{00}$ if $\rho \uparrow$ and $\uparrow \kappa_{ij}$ if $\rho \uparrow$; Right: $\uparrow \lambda_{00}$ if $\rho \uparrow$.

The case of multiple alternative hypotheses

In the local model test with a single alternative hypothesis \mathcal{H}_i of Eq. (9), the sizes of type I and II errors are given. Under this assumption, the non-centrality parameter coinciding with the MDB can be obtained as a lower bound of the outlier that can be successfully detected (Yang *et al.* 2013). In practice, however, we do not have a single alternative hypothesis during the data snooping procedure, but we have multiple alternative hypotheses. Therefore, the data snooping procedure has an effect when it returns the largest absolute value among the w_i (Teunissen 2006), i.e.

$$w = \max_{i \in \{1, \dots, n\}} |w_i| \quad (19)$$

Eq. (19) indicates that the observation that is most likely an outlier corresponds to the extreme value of Baarda's w -test. In order to check its significance, the maximum value (w) should be compared with a critical value (\sqrt{k}). For instance, a j th observation is suspected to be an outlier (i.e. we accept \mathcal{H}_j) when

$$|w_j| > |w_i| \forall i, |w_j| > \sqrt{k} = \sqrt{\chi_{\alpha_0}^2(1,0)} \quad (i \neq j) \quad (20)$$

If none of the n w -tests gets rejected, then we accept the null hypothesis \mathcal{H}_0 . The data snooping procedure for the case of multiple alternative hypotheses is therefore given as

$$\text{Accept } \mathcal{H}_0 \text{ if } w = \max_{i \in \{1, \dots, n\}} |w_i| \leq \sqrt{k} \quad (21)$$

Otherwise,

$$\text{Accept } \mathcal{H}_i \text{ if } w = \max_{i \in \{1, \dots, n\}} |w_i| > \sqrt{k} \quad (22)$$

The computational complexity increases considerably as the number of hypotheses increases (Yang *et al.* 2013). This means testing \mathcal{H}_0 against $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3, \dots,$ and \mathcal{H}_n . However, dealing with n alternative hypotheses is not a trivial task for quality control purposes, because the higher the dimensionality of the alternative hypotheses, the more complicated the level probabilities associated with the data snooping procedure.

The acceptance region for \mathcal{H}_0 is the intersection of the n individual acceptance regions, i.e. $\mathcal{A}_{\alpha_n} = \bigcap_{i=1}^n \mathcal{A}_{\alpha_0}^i, (\forall i = 1, \dots, n)$. Because the acceptance region of the multiple test \mathcal{A}_{α_n} is smaller than the single alternative hypothesis $\mathcal{A}_{\alpha_0}^i$ (Eq. 12), the probability of type II error through Eq. (21) will be

$$\beta_n = P[\mathbf{y} \in \mathcal{A}_{\alpha_n} | \mathcal{H}_i] \leq P[\mathbf{y} \in \mathcal{A}_{\alpha_0} | \mathcal{H}_i] = \beta_0 \quad (23)$$

where β_n is the type II error probability for the multiple alternative hypotheses, whereas β_0 is for the case of one alternative hypothesis (note: the index ‘0’ denotes the univariate probability, whereas n represents the multivariate probability).

Geometrically, the intersection of all n individual acceptance regions has a smaller area than each of the individual ones. Hence, the type II error probability becomes smaller, i.e. $\beta_n \leq \beta_0$. Therefore, we also have the following MDB inequality:

$$\text{MDB}_n \leq \text{MDB}_0 = \sigma_{\bar{v}_i} \sqrt{\lambda_{q=1}(\alpha_0, \gamma_0, r = q = 1)} \quad (24)$$

The right side of the inequality (24) is typically easy to compute using Eq. (16). Taking into consideration the geometric complexity of the acceptance region and the probability density function of the w -tests, the computation of the left side requires a Monte Carlo integration [see Imperato *et al.* (2018) for more details]. In multiple tests, on the other hand, for type I error, we have

$$\alpha_n = P[\mathbf{y} \notin \mathcal{A}_{\alpha_n} | \mathcal{H}_0] \geq P[\mathbf{y} \notin \mathcal{A}_{\alpha_0} | \mathcal{H}_0] = \alpha_0 \quad (25)$$

According to Eq. (23), the type I error will generally become larger, i.e. $\alpha_n \geq \alpha_0$. By using the concept of Bonferroni (1936) and neglecting the dependence between w -tests, the type I error for multiple tests can be given as $1 - \alpha_n = (1 - \alpha_0)^n$, where for small α_0 , it can be approximated as $\alpha_n \leq n\alpha_0$. As pointed out by Imparato *et al.* (2018), this approximation works very well for a system with a high redundancy and/or low correlated w -tests. Lehmann (2012) used the Monte Carlo simulation by considering the dependencies between the least-squares residuals. In that case, as n increases, the values of the correlation coefficients decrease.

In addition, a wrong exclusion is caused by a high correlation between Baarda's w -test statistics. Therefore, the separability analysis for controlling misidentification when there are multiple alternative hypotheses should also be considered. For the multiple alternative hypotheses case, the probability of committing different types of error are presented in Table 1.

Table 1. Decisions for the multiple alternative hypotheses case (adapted from Yang *et al.* 2013)

Reality (‘unknown’)	Result of test				
	\mathcal{H}_0 $ w_i \leq \sqrt{k}, \forall i$	\mathcal{H}_1 $ w_1 > \sqrt{k}, w_1 > w_i , \forall i$	\mathcal{H}_2 $ w_2 > \sqrt{k}, w_2 > w_i , \forall i$...	\mathcal{H}_n $ w_n > \sqrt{k}, w_n > w_i , \forall i$
\mathcal{H}_0	Correct decision $1 - \alpha_n$	Type I Error α_{01}	Type I Error α_{02}	...	Type I Error α_{0n}
\mathcal{H}_1	Type II Error β_{10}	Correct identification $1 - \beta_{11}$	Type III Error κ_{12}	...	Type III Error κ_{1n}
\mathcal{H}_2	Type II Error β_{20}	Type III Error κ_{21}	Correct identification $1 - \beta_{22}$...	Type III Error κ_{2n}
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
\mathcal{H}_n	Type II Error β_{n0}	Type III Error κ_{n1}	Type III Error κ_{n2}	...	Correct identification $1 - \beta_{nn}$

Under multiple alternative hypotheses, the type I errors in the data snooping procedure for outlier identification, when there are no outliers, are given by

$$\begin{aligned}\alpha_{0i} &= \text{P}[|w_i| > |w_j| \forall j, |w_i| > \sqrt{k}(j \neq i) | \mathcal{H}_0: \text{true}] \\ &= \int_{|w_i| > |w_j| \forall j, |w_i| > \sqrt{k}} f'_0 \partial w_i \dots \partial w_n\end{aligned}\quad (26)$$

where f'_0 is the probability density function when the expectation of the multivariate Baarda's w -test statistics is zero ($\mu_n = 0$).

The confidence level is

$$\begin{aligned}1 - \alpha_n &= \text{P}[\bigcap_{i=1}^n |w_i| \leq \sqrt{k} | \mathcal{H}_0: \text{true}] \\ &= \int_{|w_i| > |w_j| \forall j, |w_i| > \sqrt{k}} f'_0 \partial w_i \dots \partial w_n\end{aligned}\quad (27)$$

Based on the assumption that one outlier is in the i th position of the dataset, the probability of a correct identification is

$$\begin{aligned}1 - \beta_{ii} &= \text{P}[|w_i| > |w_j| \forall j, |w_i| > \sqrt{k}(j \neq i) | \mathcal{H}_i: \text{true}] \\ &= \int_{|w_i| > |w_j| \forall j, |w_i| > \sqrt{k}} f'_i \partial w_i \dots \partial w_n\end{aligned}\quad (28)$$

where f'_i is the probability density function when the expectation of the multivariate Baarda's w -test statistics is not equal to zero ($\mu_n \neq 0$).

The size of type II error for multiple testing is

$$\beta_{i0} = \text{P}[\bigcap_{i=1}^n |w_i| \leq \sqrt{k} | \mathcal{H}_i: \text{true}]\quad (29)$$

and the size of type III error is given by

$$\sum_{i=1}^n \kappa_{ij} = \sum_{i=1}^n \mathbf{P}[|w_j| > |w_i| \forall i, |w_j| > \sqrt{k}(j \neq i) | \mathcal{H}_i: \text{true}] \quad (30)$$

The probability levels associated with Baarda's data snooping under multiple alternative hypotheses by using an analytical formulation were described by Yang *et al.* (2013), as demonstrated above. However, these authors have already pointed out that the density functions and the critical region under multiple alternative hypotheses are extremely complex and even impossible to obtain deterministically. They used an empirical method based on numerical simulation of random observation errors to analyse their analytical expression. The relationship between the number of observations (n alternative hypotheses), the probabilities of committing different types of errors during data snooping procedure, and the non-centrality parameter for multiple alternative hypotheses (denoted by $\lambda_{n,q=1}$) are shown in Fig. 4. The results were based on the simulation performed by Yang *et al.* (2013) in which the correlation coefficients between Baarda's w -test statistics were assumed to be the same.

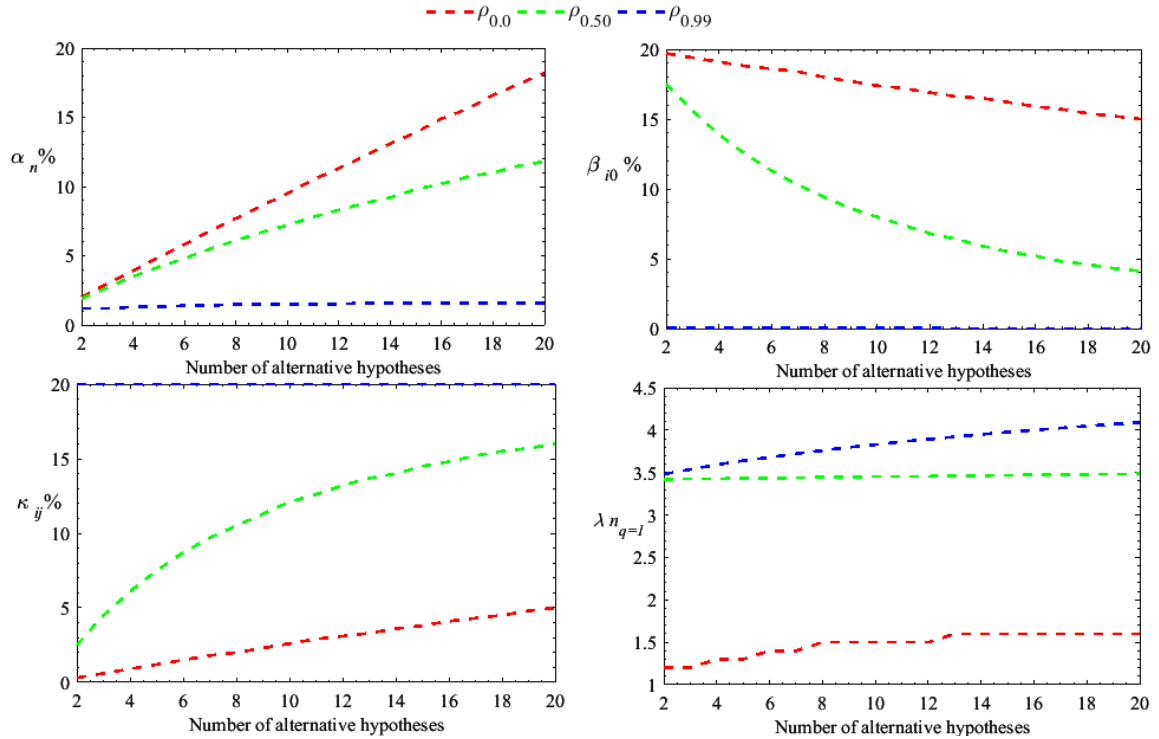


Figure 4. Results of multiple alternative hypotheses at fixed values of $\alpha_0 = 0.01$ and $1 - \beta_{ii} = 0.8$. Top left: $\uparrow \alpha_n$ ($\downarrow \sqrt{k}$) if $\rho \downarrow$ and for $\rho < 0.99$, $\uparrow \alpha_n$ ($\downarrow \sqrt{k}$) if $n \uparrow$. Top right: $\uparrow \beta_{i0}$ if $\rho \downarrow$ and for $\rho < 0.99$, $\uparrow \beta_{i0}$ ($\downarrow \sqrt{k}$) if $n \downarrow$. Bottom left: $\uparrow \kappa_{ij}$ if $\rho \uparrow$ and for $\rho < 0.99$, $\uparrow \kappa_{ij}$ if $n \uparrow$. Bottom right: $\uparrow \lambda_{n_{q=1}}$ if $\rho \uparrow$.

From Fig. 4, although α_n increases with the number of alternative hypotheses, α_n decreases significantly as the correlation coefficient ρ increases. The non-centrality parameter $\lambda_{n_{q=1}}$ notably increases as the correlation coefficient ρ increases and it also increases as the number of alternative hypotheses n increases. Thus, in order to have a low probability of committing a type II or III error, a larger $\lambda_{n_{q=1}}$ should be required. Gradually, the size of type II error diminishes as the number of alternative hypotheses increases. The larger the correlation coefficient, the lower the probability of committing β_{i0} and the higher the probability of committing type III error κ_{ij} .

In this context, Prószyński (2015) provided a supplementation of the MDB by an outlier identifiability index to each observation and a misidentifiability index as the maximum probability of type III error. Assuming that one outlier of MDB magnitude

resides in the i th position of the dataset, the identifiability index (here denoted as $ID_{(i)}$) is defined in terms of conditional probability as follows:

$$ID_{(i)} = P \left[\left(\bigcap_{j=1}^{n-1} \frac{|w_j|}{|w_i|} < 1, (j \neq i) \mid \bigcup_{j=1}^n |w_j| > \sqrt{k} \right) \right] \quad (31)$$

Moreover, the misidentifiability index ($MID_{(ij)}$) as the probability of identifying a j th non-outlying observation instead of the contaminated i th observation is

$$MID_{(ij)} = P \left[\left(\bigcap_{j=1}^{n-1} \frac{|w_i|}{|w_j|} < 1, (j \neq i) \mid \bigcup_{j=1}^n |w_i| > \sqrt{k} \right) \right] \quad (32)$$

Eq. (31) indicates that the statistic of the outlier in the i th position dominates over each of the corresponding absolute values for the remaining observations within a set of suspected observations. Eq. (32), on the other hand, shows that the statistic of the non-outlying observation in the j th position is prevailing within a set of suspected observations. Thus, the ratio between the w -statistics for both Eq. (31) and Eq. (32) is smaller than one.

Prószyński (2015) then proposed the maximum value of $MID_{(ij)}$, i.e. the maximum probability of committing type III errors. Furthermore, he formulated the following relationship based on Eqs. (31) and (32):

$$ID_{(ij)} = 1 - \sum_{j=1, j \neq i}^{n-1} MID_{(ij)} \quad (33)$$

with n as in Eqs. (31) and (32) being the number of observations (or alternative hypotheses). The indices in Eq. (33) show that the higher the probability of finding the contaminated observation (Wang and Knight 2012), the lower the probability of committing the type III error. The outliers of MDB magnitudes are often not identified. Thus, Prószyński (2015) also provided a multiplying factor, which indicates the degree of augmentation of MDB, to obtain the required level of outlier identifiability. This

multiplying factor is based on another index called partial pseudo-identifiability index ($ID_{(ij)}^*$), which is given by

$$ID_{(ij)}^* = P \left[\frac{|w_j|}{|w_i|} < 1, (j \neq i) \right] \quad (34)$$

An analogy can be found between the proposal by Prószyński (2015) and Wang and Knight (2012). In the approach by the latter, the concept of minimal separable bias (MSB), which is the magnitude of MDB increased by the multiplying factor to ensure identification of an outlier at a satisfactory confidence level, is presented. When the number of observations increases, the definitions of Eqs. (31), (32), (33), and (34) become more complex. Therefore, an empirical method based on numerical simulation of random observation errors was used to analyse the above analytical expression [for more details, see Prószyński (2015)].

In the context of identifiability of the hypotheses, Teunissen (2017) recently introduced the concept of MIB as the smallest outlier that leads to its identification for a given correct identification rate. The detection and identification are equal in the case where we only have the one alternative hypothesis. However, under n alternative hypotheses (multiple testing), we have from Eqs. (28), (29), and (30)

$$\beta_{ii} = \beta_{i0} + \sum_{i=1}^n (j \neq i) \kappa_{ij} \quad (35)$$

or

$$1 - \beta_{ii} = \gamma_0 - \sum_{i=1}^n (j \neq i) \kappa_{ij} \therefore \gamma_0 = 1 - \beta_{ii} + \sum_{i=1}^n (j \neq i) \kappa_{ij} \quad (36)$$

The probability of correct detection γ_0 (power of the test for a single alternative hypothesis) is the sum of the probability of correct identification $1 - \beta_{ii}$ (selecting a correct alternative hypothesis) and the probability of misidentification $\sum_{i=1}^n (j \neq i) \kappa_{ij}$

(selecting one of the $n - 1$ other hypotheses). Thus, we have the inequality $1 - \beta_{ii} \leq \gamma_0$ (Imparato *et al.* 2018), which is similar to Eq. (23). As a consequence of that inequality, the MIB will be larger than MDB:

$$\text{MIB}_n \geq \text{MDB}_0 \quad (37)$$

The computation of MIB_n should be based on simulations, because the acceptance region (as well as the critical region) for the multiple alternative hypotheses case is intractable and even impossible to obtain by numerical integration or by a closed analytical form. In this respect, Imparato *et al.* 2018 showed how to compute the quantities MIB_n and MDB_n . They found that the larger the size of the outlier and/or more precisely, the estimated outlier, the higher the probability of being correctly identified. In addition, increasing the type I error (i.e. reducing the acceptance region) leads to higher probabilities of correct identification. Furthermore, increasing the number of alternative hypotheses leads to a lower probability of correct identification. Therefore, the general characteristic of MIB_n , MDB_n , and MDB_0 is given as

$$\text{MIB}_n \geq \text{MDB}_0 \geq \text{MDB}_n \quad (38)$$

There is no difference between MDB and MIB in the case of a single alternative hypothesis. As the number of alternative hypotheses increases, however, MDBs become smaller, whereas MIBs become larger. The relation in (38) shows that MDB_0 can be used as a safe upper bound for its multivariate MDB_n . However, this is not the case for MIBs.

Data snooping with adaptation: an iterative procedure

From the MIB concept, one may evaluate, using an a priori analysis, the probability of identifying an outlier in the first adjustment run. In that case, the data snooping

procedure is applied only once according to the detector (Eq. (19)). In practice, however, data snooping is applied iteratively in the process of estimation, identification, and adaptation. Fig. 5 illustrates an example of data snooping applied iteratively to a regression model for the case where there are two outliers in the dataset and it is assumed that false decisions, such as types I, II, and III errors, do not occur.

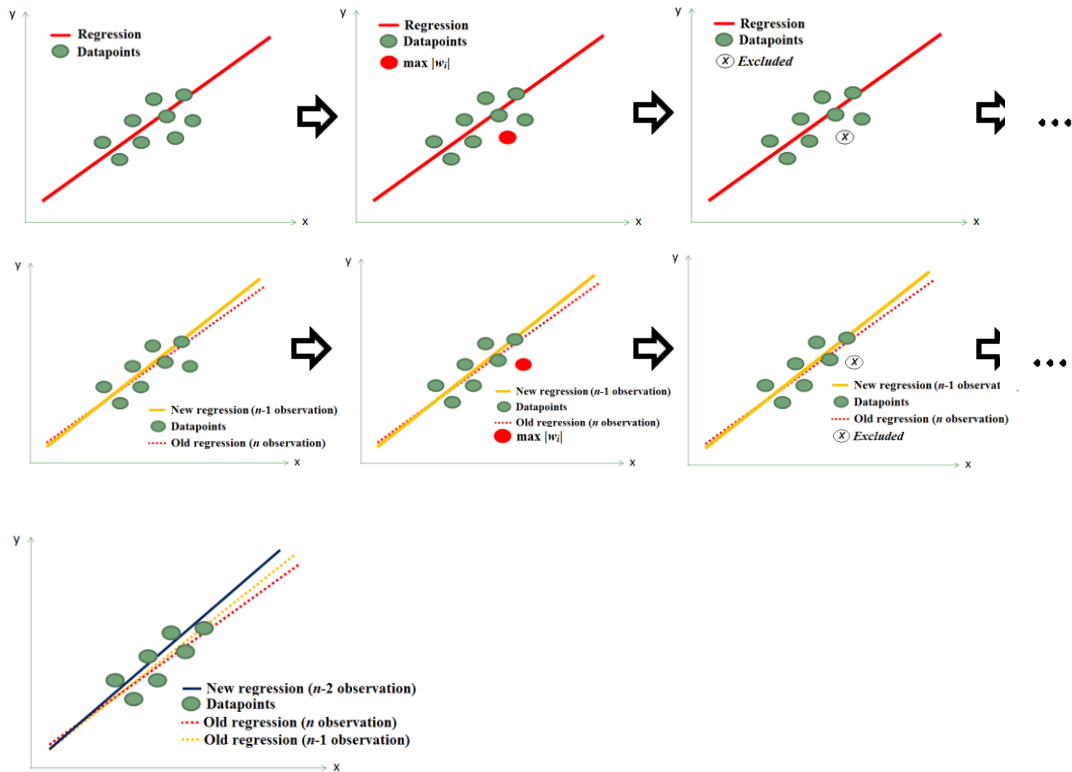


Figure 5. Data snooping procedure in its iterative form for the case where there are two outliers in the dataset.

First, the least-squares residual vector is estimated and Baarda's w -test statistics are computed by Eq. (10). Then, the detector given by Eq. (19) is applied to identify the most likely outlier (red mark in Fig. 5). The identified outlier is then excluded from the dataset (marked with 'x' in Fig. 5) and the least-squares estimation adjustment is restarted without the rejected observation. Then, Baarda's w -test (Eq. 11) as well as the detector (Eq. 19) are again computed. Obviously, if redundancy permits, this procedure

is repeated until no more (possible) outliers can be identified. This procedure is called IDS (Teunissen 2006, pp. 135).

The theories presented until the present are based on a single run of data snooping, i.e. without subsequent diagnostic operations such as removal or re-weighting of observations. However, the actual practice of data snooping is performed in its iterative process. Thus, as the case in actual practice, a reliability measure cannot be easily computed for quality control purposes. Consequently, MDB , ID , $ID_{(ij)}^*$, MID , and MIB are valid only for the case where data snooping is run once, and they cannot be used as a diagnostic tool for IDS. Because an analytical formula is not easy to compute, a simulation technique should be run to obtain the probabilities and reliability measures for IDS. Numerical simulations allow insights into these cases where analytical solutions are extremely complex to fully understand, are doubted for one reason or another, or are not available. In the next section, we present the Monte Carlo simulation method as an alternative for quality control.

Monte Carlo simulation: an alternative for quality control

The foundation of the Monte Carlo method was Buffon's needle problem by Georges Louis Leclerc in the eighteenth century. Later, in the nineteenth century, William Sealy Gosset, otherwise known as 'Student', Fisher's disciple, discovered the form of the 't-distribution' by a combination of mathematical and empirical work with random numbers, which is now known as an early application of the Monte Carlo simulation. However, the Monte Carlo simulations became well known in the 1940s, when Stanisław Ulam, Nicholas Metropolis, and John von Neumann worked on the atomic bomb project. That method was used to solve the problem of diffusion and absorption of neutrons, which was difficult to consider in any analytical approaches (Stigler 2002).

Despite advances in science and technology to solve highly complex systems, one of the major obstacles to simulation up until the 1980s was the analysis time and computing resources (run time and memory). However, the advent of personal computers with powerful processors has rendered simulation a particularly attractive and cost-effective approach to performance analysis of complex systems. Therefore, the Monte Carlo simulation emerged as a solution to help analysts understand how well a system performs under a given regime or a set of parameters.

The key of the Monte Carlo method is the random number generator. A random number generator is an algorithm that generates a deterministic sequence of numbers, which simulates a sequence of independent and identically distributed (*i.i.d.*) numbers chosen uniformly between 0 and 1. It is 'random' in the sense that the sequence of numbers generated passes the statistical tests for randomness. For this reason, random number generators are typically referred to as PRNGs. PRNGs are part of many machine learning and data mining techniques. In simulation, a PRNG is implemented as a computer algorithm in some programming language, and is made available to the user via procedure calls or icons (Altiok and Melamed 2007). A good generator produces numbers that are not distinguishable from *truly random numbers* in a limited computation time. This is, in particular, true for Mersenne Twister (Matsumoto and Nishimura 1998), a popular generator with a long period length of $2^{19937} - 1$.

Typically, the random errors of good measurements are normally distributed with expectation zero. In order to have normal random errors, uniformly distributed random number sequences (produced by the Mersenne Twister algorithm, for example) are transformed into a normal distribution by using the Box–Muller transformation (Box and Muller 1958). Box–Muller has been used in geodesy for Monte Carlo simulation (Lehmann 2012). In essence, the Monte Carlo method replaces random variables by

computer PRNGs, probabilities by relative frequencies, and expectations by arithmetic means over large sets of such numbers. A computation with one set of PRNG is a Monte Carlo experiment (Lehmann and Scheffler 2011), also referred to as the number of Monte Carlo simulations.

It is evident that in the last decades, the use of Monte Carlo technique for quality control proposals in geodesy has been increasing. Hekimoglu and Koch (1999) pioneered the idea of using simulation techniques to geodesy for evaluating some probabilities as simple ratios from simulated experiments. Lehmann and Scheffler (2011) used 5000 Monte Carlo experiments to obtain type I errors such that the desired parameters are best estimated. Aydin (2012) used 5000 simulations to investigate the global test procedure in deformation analysis. Yang *et al.* (2013) used simulations to analyse the probability levels of data snooping. Koch (2016) investigated the non-centrality parameter of the F-distribution by using 100,000 simulated random variables. Prószynski (2015) showed that the degree of dispersion of the *ID* values by 1000 simulations can reach ± 1 or $\pm 2\%$ (standard deviations). Klein *et al.* (2017) ran 1000 experiments to verify the performance of sequential likelihood ratio tests for multiple outliers. Lehmann and Voß-Böhme (2017) dealt with the generalised chi-squared distribution χ^2 by using the Monte Carlo method and concluded that it is possible to outperform the statistical power of Baarda's *w*-test. Recently, Imperato *et al.* (2018) used 100,000 Monte Carlo simulations to analyse the differences between MDB and MIB, as those quality control measures are not easy to compute for multiple alternative hypotheses, as already shown here.

Despite countless advances in quality control mainly because of the increase in the use of computer simulation techniques, there is continued research on the topic, which paves the way for further investigations. In this section, we introduce an

approach for evaluating the accuracy of Monte Carlo simulation. In addition, we also show how to obtain the probability levels associated with Baarda's IDS procedure by Monte Carlo simulation. In the latter case, an example based on a simulated levelling network is considered for the analysis.

Accuracy of Monte Carlo simulation for quality control purposes

The larger the number of simulations (i.e. the sample size of random numbers generated), the better the approximation of an expected value in some stochastic processes. The theoretical convergence is only of the order of $m^{1/2}$ (where m is the number of simulations performed). This order does not depend on the dimension of the dataset (Tanizaki 2004). However, until the present, no study has been conducted to evaluate empirically the accuracy of the Monte Carlo simulation for quality control purposes in geodesy. Generally, only the degree of dispersion of the Monte Carlo simulation technique is considered. Thus, an issue remains: how can we find the optimal number of Monte Carlo simulations for quality control purpose? We use an exact theoretical reference given by a univariate minimal detectable bias MDB_0 (Eq. 14).

The procedure to analyse the accuracy of Monte Carlo simulation can be briefly described as follows:

- (1) Generate a vector of random errors. Typically, it is assumed that the random errors of the good measurements are normally distributed with expectation zero. Thus, the random errors are generated by using the well-known Box–Muller method (Box and Muller 1958) based on a multivariate normal distribution of the observations (actually, we used MATLAB's function *mvnrnd* here).
- (2) Add the MDB_0 computed for a given observation using the analytical expression (Eq. 14) to the corresponding random error. Thus, we have the total error (here,

denoted by $\boldsymbol{\varepsilon}$) as a combination of the random errors and the observation contaminated by its corresponding MDB₀, i.e.

$$\boldsymbol{\varepsilon} = \mathbf{e} + \mathbf{c}_i \text{MDB}_{0(i)} \quad (39)$$

where $\mathbf{e} \in \mathbb{R}^n$ is the random error generated from normal distribution $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \Sigma_{yy})$ and \mathbf{c}_i consists exclusively of elements with values of 0 and 1, where 1 means that an i th outlier of magnitude $\text{MDB}_{0(i)}$ affects an i th measurement, and 0 otherwise.

(3) Compute the least-squares residual $\hat{\mathbf{e}}_0$ vector as follows:

$$\hat{\mathbf{e}}_0 = \mathbf{R}\boldsymbol{\varepsilon}, \text{ with } \mathbf{R} = \mathbf{I} - \mathbf{A}(\mathbf{A}^\top \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{W} \quad (40)$$

where $\mathbf{R} \in \mathbb{R}^{n \times n}$ is the redundancy matrix, \mathbf{W} is the weight matrix, and $\mathbf{I} \in \mathbb{R}^{n \times n}$ is the identity matrix.

(4) Compute Baarda's w -statistics given by Eq. (10). If $|w_i| > \sqrt{\chi_{\alpha_0}^2(r=1,0)} = \sqrt{k}$, then \mathcal{H}_0 is rejected in favour of \mathcal{H}_i .

(5) Perform the procedure described from 1 to 4 above for m samples of random error vector \mathbf{e}_v with $v = \{1, \dots, m\}$. By counting the number of correct detections in the generated samples, one can determine the corresponding power of the test. Thus, if m is the total number of samples simulated, we count the number of times of correct detection n_{CD} in which $|w_i^v| > \sqrt{\chi_{\alpha_0}^2(r=1,0)} = \sqrt{k}$ for $v = \{1, \dots, m\}$, and then approximate the probability of the power of the test as

$$\hat{\gamma}_0 = \frac{n_{CD}}{m} \quad (41)$$

where m is known as the number of Monte Carlo simulations (or also referred to as

Monte Carlo sample size).

One of the potential disadvantages of the Monte Carlo simulation is that a single run of a Monte Carlo sample does not indicate by itself the reliability of the results. For analysing the accuracy of Monte Carlo method, several trials for each sample size should be performed. By computing the average and standard deviations of the trials, one can analyse the accuracy as a function of the Monte Carlo sample size. The average empirical probability of the power of the test $\bar{\gamma}_0$ should be as close as possible to the true value (i.e. γ_0). The average is given by

$$\bar{\gamma}_0 = \frac{1}{t} \sum_{i=1}^t \hat{\gamma}_{0(i)} \quad (42)$$

where $\hat{\gamma}_{0(i)}$ is the empirical power of the test computed for each trial i and for a given Monte Carlo sample size m , and t is the total number of trials. The variance can also be computed to evaluate the degree of variation of the Monte Carlo sample size as follows:

$$\bar{s}_{\bar{\gamma}_0}^2 = \frac{1}{t-1} \sum_{i=1}^t (\hat{\gamma}_{0(i)} - \bar{\gamma}_0)^2, \text{ for } i = 1, \dots, t \quad (43)$$

with $\bar{s}_{\bar{\gamma}_0}^2$ being the variance and $\bar{s}_{\bar{\gamma}_0} = \pm \bar{s}_{\bar{\gamma}_0}^{1/2}$ the standard deviation for a given Monte Carlo sample size.

For simplicity, a simulated GNSS (Global Navigation Satellite System) network is considered. It has one control station (fixed) and five user stations with unknown 3D positions (X, Y, Z), totalling six minimally constrained stations. For each pair of stations, there are four of five baselines (see Klein *et al.* (2017) for more details of that geodetic network). Here, we choose a significance level of $\alpha_0 = 0.001$ and a power of the test $\gamma_0 = 0.8$. Based on those probabilities, the MDB was computed for the baseline MGIN-SPCA of that geodetic network (see Eq. (14)). We run 1.10^3 , 5.10^3 , 1.10^4 , 2.10^4 , 3.10^4 , 4.10^4 , 5.10^4 , 6.10^4 , 7.10^4 , 8.10^4 , 9.10^4 , and 10.10^4 Monte Carlo sample sizes. For

each Monte Carlo sample size, 1000 trials were performed to evaluate the Monte Carlo simulations.

Table 2 presents the statistics of the power of the test computed by simulations as well as the mean elapsed time for each Monte Carlo sample size. In our experiments, Intel Core i5-4200M and Matlab (R2017a) were used for the simulations. Fig. 6 shows the variation of the empirical power of the test computed by simulation for each Monte Carlo sample size.

Table 2. Statistics of Monte Carlo simulation based on computation of the power of the test for 1000 trials.

Number of Monte Carlo simulations	$\bar{\gamma}_0 (\pm \bar{s}_{\bar{\gamma}_0})$ (%)	Min. $\hat{\gamma}_{0(i)}$ (%)	Max. $\hat{\gamma}_{0(i)}$ (%)	Mean elapsed time (s)
$1 \cdot 10^3$	80.02 (± 1.31)	75.10	83.80	$5 \cdot 10^{-2}$
$5 \cdot 10^3$	80.01 (± 0.56)	77.98	81.60	$2 \cdot 10^{-1}$
$1 \cdot 10^4$	80.02 (± 0.39)	78.80	81.21	$5 \cdot 10^{-1}$
$2 \cdot 10^4$	80.01 (± 0.28)	79.12	80.77	$9 \cdot 10^{-1}$
$3 \cdot 10^4$	79.99 (± 0.23)	79.26	80.65	1.3
$4 \cdot 10^4$	80.00 (± 0.21)	79.29	80.65	1.8
$5 \cdot 10^4$	79.99 (± 0.18)	79.38	80.53	2.2
$6 \cdot 10^4$	80.00 (± 0.16)	79.53	80.48	2.6
$7 \cdot 10^4$	79.99 (± 0.16)	79.49	80.59	3.0
$8 \cdot 10^4$	79.99 (± 0.14)	79.57	80.36	3.4
$9 \cdot 10^4$	79.99 (± 0.13)	79.59	80.40	3.7
$10 \cdot 10^4$	80.00 (± 0.13)	79.55	80.41	4.6

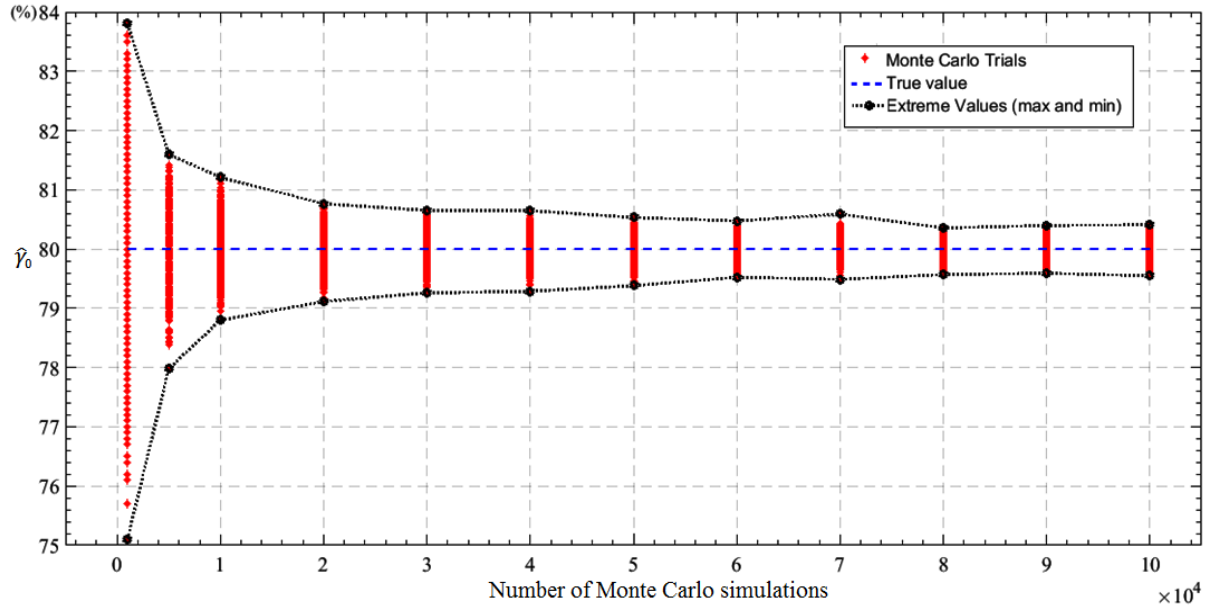


Figure 6. Variation of the empirical power of the test $\hat{\gamma}_0$ for each Monte Carlo simulation for 1000 trials. The blue horizontal dashed line at $\hat{\gamma}_0 = 80\%$ is the true value of the power of the test used for calculating MDB ($\gamma_0 = 0.8$). The black dotted line shows the maximum and minimum values of $\hat{\gamma}_0$. As expected, a small number of Monte Carlo samples (i.e. number of Monte Carlo simulations) provide a wider spread of results.

According to Table 2 and Fig. 6, the greater the number of Monte Carlo samples, the more stable the standard deviation of the power of the test $\bar{s}_{\hat{\gamma}_0}$. This property of the Monte Carlo simulation method can thus be used as a direct way of determining the number of simulations (i.e. Monte Carlo sample size) for a given application. For the application of outlier identification, which is the topic in the next section, we suggest that the Monte Carlo sample size should be chosen when $\hat{\gamma}_0$ in Eq. (41) fluctuates by no more than 0.1%. It should be noted that the tested simulations are not sufficient to compute the probabilities of order 10^{-3} .

To ensure that the approximate value of $\hat{\gamma}_0$ is close to the true value γ_0 in the order of 0.1%, we also ran for $2 \cdot 10^5$, $5 \cdot 10^5$, and 10^6 Monte Carlo simulations. The latter two simulations provided a degree of dispersion of the order of 0.06% and 0.04%, respectively. For the case of $2 \cdot 10^5$ simulations, we obtained a standard deviation of

$\pm 0.07\%$. Both $5 \cdot 10^5$ and 10^6 Monte Carlo sample sizes did not present significant advantages over the 200,000 simulations. Therefore, we consider that the number of simulations of 200,000 can provide consistent results with sufficient numerical precision for outlier identification.

Probabilities of IDS based on Monte Carlo simulation

The probability of identifying an outlier is still a bottleneck in geodesy. The challenge is even greater for the case of IDS, i.e. when data snooping is applied iteratively.

Unfortunately, the probability levels associated with IDS are not easy to study using analytical models owing to the paucity or lack of practically computable solutions (closed form or numerical). In contrast, a simulation model can almost always be constructed and run to generate system histories that yield useful statistical information on system operation and performance measures (Altiok and Melamed 2007, Robert and Casella 2013, Gamerman and Lopes 2006).

Recent studies by Rofatto *et al.* (2017) showed how to extract the probability levels associated with Baarda's IDS procedure by Monte Carlo simulation. Furthermore, they introduced two new classes of wrong decisions for IDS, which they called over-identification. One is the probability of IDS flagging simultaneously the outlier and good observations. Second is the probability of IDS flagging only the good observations as outliers (more than one) while the outlier remains in the dataset. Obviously, these two new false decisions could occur during the iterative process of estimation, identification, and exclusion, as is the case of IDS. However, more studies on the subject are still required. They have considered 10,000 simulations, which is not ideal as shown in the previous section.

A procedure based on the Monte Carlo simulation is applied to compute the probability levels of IDS as follows (summarised as a flowchart in Fig. 7). In the first

step, the design matrix $A \in \mathbb{R}^{n \times u}$ and the covariance matrix of the measurements $\Sigma_{yy} \in \mathbb{R}^{n \times n}$ are entered; then, the significance level (α_n) and the magnitude intervals of simulated outliers are defined. The magnitude intervals of outliers are based on a standard deviation of observation (e.g. $|3\sigma$ to $9\sigma|$, where σ is the standard deviation of observation). The random error vectors are synthetically generated based on a multivariate normal distribution, because the assumed stochastic model for random errors is based on a matrix covariance of the observations. Here, we use the Mersenne Twister algorithm to generate a sequence of random numbers and Box–Muller to transform it into a normal distribution. On the other hand, the magnitude of the outlier (for $q = 1$) is selected based on magnitude intervals of the outliers for each Monte Carlo simulation. We use the continuous uniform distribution to select the outlier magnitude. Thus, similar to the previous section, the total error ($\boldsymbol{\varepsilon}$) is a combination of the random errors, and its corresponding outlier is as follows:

$$\boldsymbol{\varepsilon} = \boldsymbol{e} + \boldsymbol{c}_i \nabla_i \quad (43)$$

where $\boldsymbol{e} \in \mathbb{R}^n$ is the random error generated from normal distribution $\boldsymbol{e} \sim \boldsymbol{N}(\mathbf{0}, \Sigma_{yy})$ and \boldsymbol{c}_i consists exclusively of elements with values of 0 and 1, where 1 means that an i th outlier of magnitude ∇_i affects an i th measurement, and 0 otherwise. Then, the least-squares residuals vector $\hat{\boldsymbol{e}}_0$ is computed similarly using Eq. (40).

For IDS, the hypothesis of Eq. (9) for $q = 1$ (one outlier) is assumed and the corresponding test statistic is computed according to Eq. (10). Then, the maximum test statistic value is computed according to Eq. (19). After identifying the observation suspected as the most likely outlier, it is typically excluded from the model, and least-squares estimation and data snooping are applied iteratively until there are no further

outliers identified in the dataset. The procedure should be performed for m samples of random error vectors with each sample contaminated by an outlier.

By counting the number of correct identifications in the generated samples, one can determine the corresponding success rate. Thus, if m is the total number of samples simulated, we count the number of times that the outlier is correctly identified (denoted as n_{CI}) so that $\max|w_i^v| > \sqrt{\chi_{\alpha_n}^2(r=1,0)} = \sqrt{k}$ for $v = \{1, \dots, m\}$, and then approximate the probability of correct identification (P_{CI}) as

$$P_{CI} \approx \frac{n_{CI}}{m} \quad (44)$$

As mentioned in the previous section, m is known as the number of Monte Carlo simulations (also referred to as Monte Carlo sample size). In addition, the misidentification probabilities are also approximated as

$$P_{MD} \approx \frac{n_{MD}}{m} \quad (45)$$

$$P_{WE} \approx \frac{n_{WE}}{m} \quad (46)$$

$$P_{over+} \approx \frac{n_{over+}}{m} \quad (47)$$

$$P_{over-} \approx \frac{n_{over-}}{m} \quad (48)$$

where n_{MD} is the number of simulations in which the IDS does not detect the outlier (P_{MD} represents the type II error, also referred to as missed detection probability); n_{WE} is the number of simulations in which the IDS procedure flags a non-outlying observation while the ‘true’ outlier remains in the dataset (P_{WE} is the probability of committing a type III error, also called wrong exclusion); n_{over+} is the number of

simulations where the IDS identifies correctly the outlying observation and others, and P_{over+} corresponds to its probability; finally n_{over-} represents the number of simulations where the IDS identifies more than one non-outlying observation, whereas the ‘true outlier’ remains in the dataset, and P_{over-} is the probability corresponding to that misidentification class.

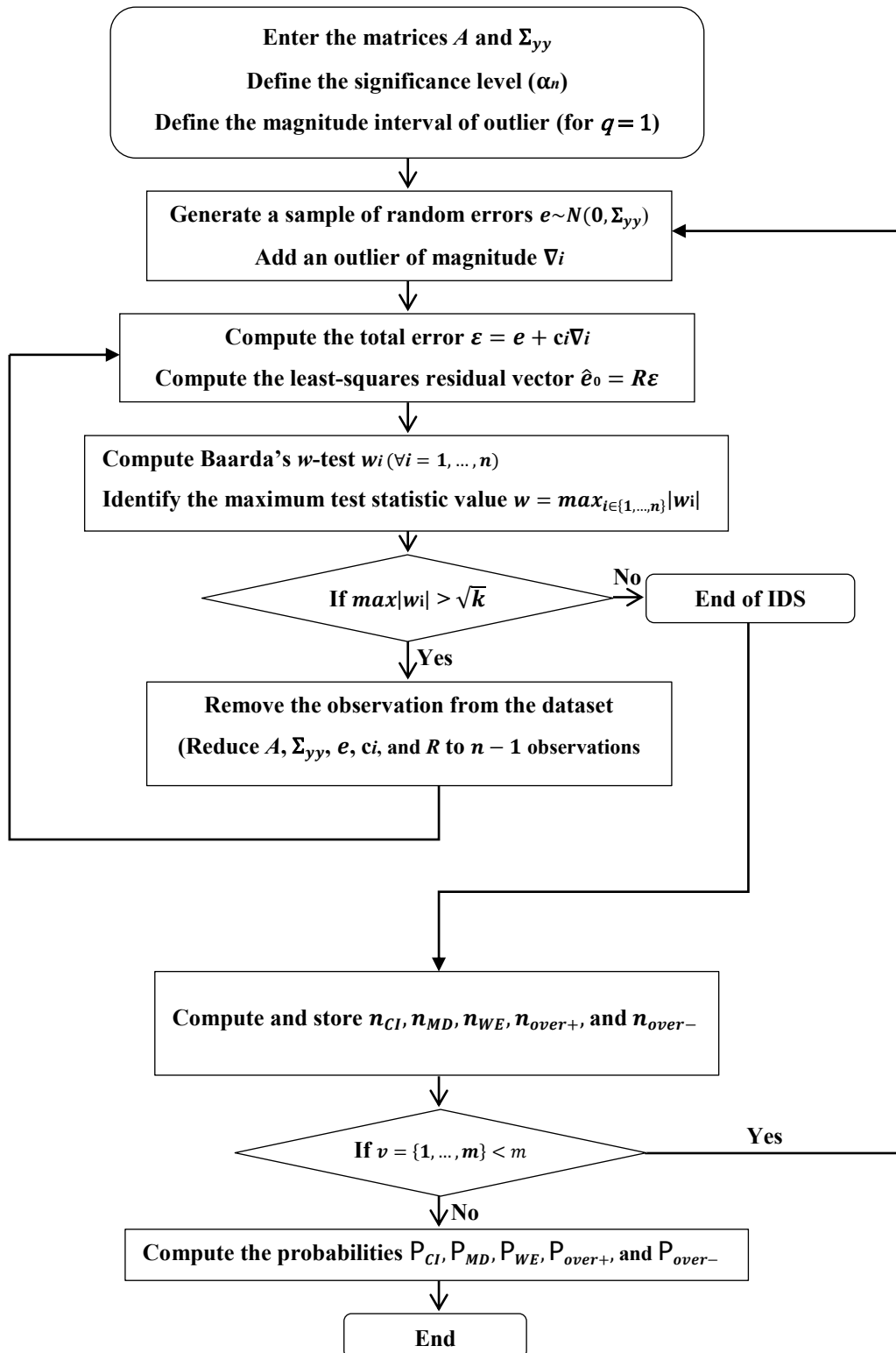


Figure 7. Flowchart of the procedure based on Monte Carlo simulation for computation of the probability levels of IDS for each observation.

As an example, the procedure based on Monte Carlo simulation for the computation of probability levels of IDS is applied to the following simulated closed-levelling network, with one control (fixed) point (A) and three points with unknown heights (B, C, and D), totalling four minimally constrained points (see Fig. 8). Thus, there are $n = 6$ observations, $u = 3$ unknowns, and $n - u = 6 - 3 = 3$ redundant observations in this simulation. Observations 1, 2, 3, 4, 5, and 6 are assumed normally distributed, uncorrelated, and with nominal precisions (a priori standard deviation σ) of ± 8 mm, ± 5.6 mm, ± 5.6 mm, ± 8 mm, ± 5.6 mm, and ± 8 mm, respectively. The magnitude interval of outlier is from the minimum 3σ to maximum 6σ , with an interval rate of 0.5σ . Here, positive and negative outliers are considered for each observation. The significance level is taken as $\alpha_n = 0.001$. We ran 200,000 Monte Carlo simulations for each observation and for each outlier magnitude interval, totalling 7,200,000 simulations. Tables 3–8 presents the correct identification and misidentification rates for one outlier and for each observation.

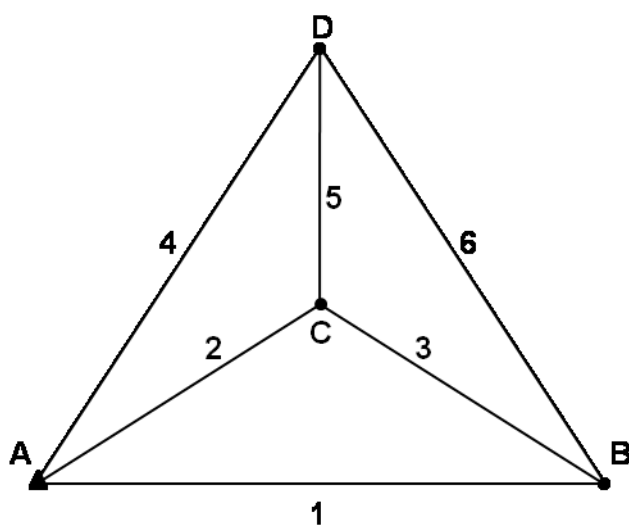


Figure 8. Simulated levelling network.

Table 3. Success and misidentification rates for one outlier in observation 1 ($\alpha_n = 0.001$).

Outlier magnitude interval	P_{CI} %	P_{MD} %	P_{over+} %	P_{over-} %	P_{WE} %
3σ to 3.5σ	21.20	75.22	0.01	0.01	3.55
3.5σ to 4σ	33.45	62.24	0.02	0.01	4.29
4σ to 4.5σ	47.95	47.54	0.02	0.01	4.48
4.5σ to 5σ	62.69	32.95	0.03	0.01	4.32
5σ to 5.5σ	75.59	20.60	0.05	0.02	3.75
5.5σ to 6σ	85.45	11.48	0.06	0.02	2.98

Table 4. Success and misidentification rates for one outlier in observation 2 ($\alpha_n = 0.001$).

Outlier magnitude interval	P_{CI} %	P_{MD} %	P_{over+} %	P_{over-} %	P_{WE} %
3σ to 3.5σ	9.91	86.69	0.00	0.00	3.40
3.5σ to 4σ	16.27	79.43	0.00	0.01	4.29
4σ to 4.5σ	25.06	69.92	0.00	0.00	5.01
4.5σ to 5σ	35.69	58.65	0.01	0.01	5.65
5σ to 5.5σ	47.39	46.68	0.01	0.01	5.90
5.5σ to 6σ	59.42	34.73	0.03	0.01	5.81

Table 5. Success and misidentification rates for one outlier in observation 3 ($\alpha_n = 0.001$).

Outlier magnitude interval	P_{CI} %	P_{MD} %	P_{over+} %	P_{over-} %	P_{WE} %
3σ to 3.5σ	9.83	86.84	0.00	0.00	3.33
3.5σ to 4σ	16.57	79.21	0.00	0.00	4.21
4σ to 4.5σ	25.11	69.69	0.01	0.01	5.19
4.5σ to 5σ	35.47	58.87	0.01	0.01	5.65
5σ to 5.5σ	47.55	46.49	0.01	0.01	5.94
5.5σ to 6σ	59.46	34.75	0.02	0.01	5.75

Table 6. Success and misidentification rates for one outlier in observation 4 ($\alpha_n = 0.001$).

Outlier magnitude interval	P_{CI} %	P_{MD} %	P_{over+} %	P_{over-} %	P_{WE} %
3σ to 3.5σ	21.04	75.32	0.01	0.00	3.63
3.5σ to 4σ	33.59	62.13	0.01	0.01	4.25
4σ to 4.5σ	48.01	47.47	0.02	0.01	4.49
4.5σ to 5σ	62.79	32.81	0.03	0.01	4.36
5σ to 5.5σ	75.63	20.53	0.06	0.01	3.77
5.5σ to 6σ	85.46	11.42	0.06	0.02	3.04

Table 7. Success and misidentification rates for one outlier in observation 5 ($\alpha_n = 0.001$).

Outlier magnitude interval	P_{CI} %	P_{MD} %	P_{over+} %	P_{over-} %	P_{WE} %
3 σ to 3.5 σ	9.82	86.81	0.00	0.00	3.36
3.5 σ to 4 σ	16.30	79.51	0.00	0.00	4.19
4 σ to 4.5 σ	25.06	69.97	0.01	0.00	4.96
4.5 σ to 5 σ	35.60	58.83	0.01	0.01	5.55
5 σ to 5.5 σ	47.45	46.56	0.02	0.01	5.96
5.5 σ to 6 σ	59.32	34.79	0.03	0.01	5.84

Table 8. Success and misidentification rates for one outlier in observation 6 ($\alpha_n = 0.001$).

Outlier magnitude interval	P_{CI} %	P_{MD} %	P_{over+} %	P_{over-} %	P_{WE} %
3 σ to 3.5 σ	21.00	75.39	0.01	0.00	3.60
3.5 σ to 4 σ	33.74	62.11	0.01	0.01	4.14
4 σ to 4.5 σ	48.11	47.38	0.02	0.01	4.48
4.5 σ to 5 σ	62.76	32.87	0.03	0.01	4.31
5 σ to 5.5 σ	75.56	20.56	0.04	0.02	3.82
5.5 σ to 6 σ	85.47	11.44	0.06	0.02	3.02

Tables 3–8 indicate that, in general, the greater the magnitude of outliers, the greater is the efficiency of IDS. In practice, as the magnitudes of outliers are unknown, one can define the probability of the correct identification in order to find the MIB of IDS for a given application. For the above geodetic network, if the probability of correct identification is taken as 0.85 (85%) and the type I error as 0.001 (0.1%), the MIB for observations 1, 4, and 6 is approximately 5.75 σ , whereas it is 7.25 σ for observations 2, 3, and 5. It will also be dependent on the functional model (i.e. geometric configuration of the network).

Regarding the cases of misidentification rates, in general, an increase in the magnitude interval of outliers leads to a decrease in the missed detection rate (type II error). Thus, it is important to note that over-identification cases are practically absent for $\alpha_n = 0.001$. However, as shown by Rofatto *et al.* (2017), for a GNSS network, the probability of committing over-identification during the IDS depends more on the

critical value than the outlier magnitude for a one-dimensional identification (one outlier). Furthermore, note that observations 1, 4, and 6 have the same behaviour of wrong exclusion rate (type III error). From 3σ to 4.5σ , there is an increase in wrong exclusion rate, but there is a decrease from 4.5σ to 6σ . Similarly, observations 2, 3, and 5 have an increase in wrong exclusion rate from 3σ to 5.5σ and a decrease from 5.5σ onwards. Based on Eq. (17), we found that the correlations between the statistics of observations 1, 4, and 6 have the same order of $\rho = 0.3289$, whereas observations 2, 3, and 5 have the same value of $\rho = 0.5$. Therefore, as expected, the behaviour of type III error is associated with the correlation coefficient. This example shows how to compute MIB for the IDS case based on the Monte Carlo simulation method. Obviously, MIB should be computed for a given probability of correct identification and significance level.

Final remarks

In this study, we provided an overview of the latest advances in the reliability theory for geodesy. We have shown that over the course of 50 years, several scientific concepts have been developed according to Baarda's initial ideas. Furthermore, we analysed the pioneering work of Baarda and the recent studies on the subject. Then, we pointed out new possibilities and perspectives on the issue, with a particular emphasis on the Monte Carlo simulation.

We highlighted that Monte Carlo simulation is a primary tool for deriving solutions to complex (or analytically intractable) problems. Although its history dates back to the eighteenth century, it came into its modern form in the motivation to develop nuclear weapons during World War II. However, up until the 1980s, simulation methods were restricted to scientists and engineers, who had supercomputers at their

disposal. However, the advent of personal computers with powerful processors has rendered simulation a particularly attractive and cost-effective approach for the performance analysis of complex systems.

Here, two problems were presented. First, how can we find the optimal number of Monte Carlo experiments for quality control purpose? We used the closed form of MDB given by Baarda (1968) to answer that question. We showed that the number of simulations of 200,000 could provide consistent results with sufficient numerical precision for outlier identification. Second, how do we obtain the probability levels of the IDS procedure? In this study, we used the Monte Carlo simulation as a key tool for studying the IDS procedure. We highlighted that, in contrast to the well-defined theories of reliability, the IDS procedure is a heuristic method, and therefore, there is no theoretical reliability measure for it. Hence, an analytical model with tractable solution is unknown, and therefore, one needs to resort to Monte Carlo simulations. Based on the work by Rofatto *et al.* (2017), we showed how to find the probability levels associated with IDS and how to obtain its MIB for each observation by means of the Monte Carlo simulation method for a given correct identification probability and significance level.

For future research, we intend to analyse the internal and external reliability of the IDS procedure as a function of the following parameters: probability levels associated to wrong decisions, number of alternative hypotheses considered, critical value adopted for the test, correlation coefficient between test statistics, and the non-centrality parameter. Then, we will try to develop a user-friendly reliability model for IDS.

References

- Alma, Ö. G., Kurt, S., and Ugur, A., 2011. Genetic algorithms for outlier detection in multiple regression with different information criteria. *Journal of Statistical Computation and Simulation*, 81(1), 29-47 (ISSN 0094-9655).
- Altiok, T. and Melamed, B. 2007. *Simulation Modeling and Analysis with Arena*, 1st ed. Academic Press (ISBN 9780123705235).
- Arnold, S. 1981. *The theory of linear models and multivariate analysis*, 1st ed. New York: Wiley.
- Aydin, C. and Demirel, H. 2005. Computation of Baarda's lower bound of the non-centrality parameter. *J Geodesy*, 78, 437-441. (<https://doi.org/10.1007/s00190-004-0406-1>).
- Aydin, C. 2012. Power of Global Test in Deformation Analysis. *J Surv. Eng.*, 138, 51-56.
- Baarda, W. 1967. Statistical concepts in geodesy. Netherlands Geodetic Commission, Publ. on geodesy, New series 2(4).
- Baarda, W. 1968. A testing procedure for use in geodetic networks. Netherlands Geodetic Commission, Publ. on geodesy, New Series 2(5).
- Baselga, S. 2011. Nonexistence of Rigorous Tests for Multiple Outlier Detection in Least-Squares Adjustment. *J Surv. Eng.*, 137, 109-112.
- Bonferroni, C.E. 1936. *Teoria statistica delle classi e calcolo delle probabilita*. Pubblicazioni del R Istituto Superiore de Scienze Economiche e Commerciali de Firenze, 8:3-62.
- Box, G.E. and Muller, M.E. 1958. A note on the generation of random normal deviates. *The Analls of Mathematical Statistics*, 29(2), 610-611.
- Ding, X. and Coleman, R. 1996. Multiple outlier detection by evaluating redundancy contributions of observations. *J Geodesy*, 70,489-498.
- Förstner, W. 1983. Reliability and discernability of extended Gauss-Markov models. In: *Seminar on mathematical models to outliers and systematic errors*, Deutsche Geodätische Kommission, Series A, no. 98. Munich, Germany, 79-103 (ISSN 978376681802).
- Gamerman, D. and Lopes, H.F. 2006. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd ed. Chapman & Hall/CRC Texts in Statistical Science.
- Ghilani, C.D. 2010. *Adjustment computations: spatial data analysis*. 5th ed. New Jersey: John Wiley & Sons.

Gui, Q., *et al.*, 2011. A Bayesian unmasking method for locating multiple gross errors based on posterior probabilities of classification variables. *J Geodesy*, 85,191-203.

Hawkins, D.M. 1980. Identification of outliers. London: Chapman & Hall.

Hekimoglu, S. and Koch, K.R. 1999. How can reliability of the robust methods be measured? In: M.O. Altan and L. Gründig, ed. Third Turkish-German joint geodetic days, 179-196.

Imparato, D., Teunissen, P.J.G. and Tiberius, C.C.J.M. 2018. Minimal Detectable and Identifiable Biases for quality control. *Survey Review (Latest Articles)*
doi:10.1080/00396265.2018.1437947

Kargoll, B. 2007. On the theory and application of model misspecification tests in geodesy. Dissertation. Universitäts- und Landesbibliothek Bonn.

Kargoll, B. 2012. On the theory and application of model misspecification tests in geodesy, Series C, vol. 674, German Geodetic Commission, Munich, Germany

Klein, I., Matsuoka, M.T., Guzzato, M.P., Souza, S.F. and Veronez, M.R. 2015. On evaluation of different methods for quality control of correlated observations. *Survey Review*, 47, 28-35.

Klein, I., Matsuoka, M.T., Guzzato, M.P. and Nievinski, F.G. 2017. An approach to identify multiple outliers based on sequential likelihood ratio tests. *Survey Review*, 49, 449-457.

Knight, N.L., Wang, J. and Rizos, C. 2010. Generalised measures of reliability for multiple outliers. *J Geodesy*, 84, 625-635.

Koch, I.E., *et al.*, 2017. Least trimmed squares estimator with redundancy constraint for outlier detection in GNSS networks. *Expert Systems With Applications*, 88, 230-237.

Koch, K.R. 1999. Parameter estimation and hypothesis testing in linear models. 2nd ed. Berlin: Springer.

Koch, K.R. 2007. Introduction to Bayesian statistics, 2nd ed. Berlin: Springer.

Koch, K.R. 2016. Minimal detectable outliers as measures of reliability. *J Geodesy* 89, 483-490.

Kok, J.J. 1982. Statistical analysis of deformation problems using Baarda's testing procedures. In: "Forty Years of Thought" Anniversary Volume on the Occasion of Prof Baarda's 65th Birthday, Delft vol 2, 470-488.

Kok, J.J. 1984. On data snooping and multiple outlier testing. US Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Charting and Geodetic Services

Lehmann, R. 1994. Adjustment in non-linear models by means of the adaptive Monte-Carlo-Integration. Allgemeine Vermessungsnachrichten, vol 7/1994. Herbert Wichmann Verlag GmbH Heidelberg (in German)

Lehmann, R. and Scheffler, T. 2011. Monte Carlo based data snooping with application to a geodetic network. J Appl. Geod., 5, 123-134.

Lehmann, R. 2012. Improved critical values for extreme normalized and studentized residuals in Gauss-Markov models. J Geodesy, 86, 1137-1146.

Lehmann, R. 2013. On the formulation of the alternative hypothesis for geodetic outlier detection. J Geodesy, 87,373-386.

Lehmann, R. 2015. Observation error model selection by information criteria vs. normality testing. Stud. Geophys. Geod., 59,489-504.

Lehmann, R. and Lösler, M. 2016. Multiple Outlier Detection: Hypothesis Tests versus Model Selection by Information Criteria. J. Surv. Eng. 142(4), (doi: 10.1061/(ASCE)SU.1943-5428.0000189)

Lehmann, R. and Voß-Böhme, A. 2017. On the statistical power of Baarda's outlier test and some alternative. J. Geod. Sci., 7,68-78.

Li, D.R. 1986. Trennbarkeit und Zuverlässigkeit bei zwei verschiedenen Alternativhypothesen im Gauß-Markoff-Modell, 3. Zeitschrift für Vermessungswesen, Heft

Marx, C. 2015. Outlier Detection by means of Monte Carlo Estimation including resistant Scale Estimation. Journal of Applied Geodesy, 9,123-142.

Matsumoto, M. and Nishimura, T. 1998. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Transactions on Modeling and Computer Simulation, 8, 3-30.

Pelzer, H. 1983. Detection of errors in the functional adjustment model. Deutsche Geodatische Kommission Seminar on mathematical models of geodetic photogrammetric point determination with regard to outliers and systematic errors, 61-70, Germany, ISSN: 9783769681802.

Prószyński, W. 1994. Criteria for internal reliability of linear least squares models. Bull. Géod., 68,161-167.

Prószyński, W. 2010. Another approach to reliability measures for systems with correlated observations. J Geodesy, 84,547-556.

Prószyński, W. 2015. Revisiting Baarda's concept of minimal detectable bias with regard to outlier identifiability. J Geodesy, 89, 993-1003.

Robert, C. and Casella, G. 2013. Monte Carlo statistical methods. Berlin: Springer.

- Rofatto, V.F., Matsuoka, M.T. and Klein, I. 2017. An Attempt to Analyse Baarda's Iterative Data Snooping Procedure based on Monte Carlo Simulation. *South African Journal of Geomatics*, 6, 416-435.
- Schaffrin, B. 1997. Reliability measures for correlated observations. *J. Eng. Surv.*, 123,126-137.
- Stigler, S.M. 2002. *Statistics on the Table: the History of Statistical Concepts and Methods*. Cambridge, Massachusetts: Harvard University Press.
- Teunissen, P.J.G. 1989. Quality control in integrated navigation systems. *IEEE Aerosp. Electron. Syst. Mag.*, 5(7),35-41.
- Teunissen, P.J.G. 1998. Minimal detectable biases of GPS data. *J Geodesy*, 72(4), 236-244.
- Teunissen, P.J.G. 2006. *Testing theory: an introduction*. 2nd ed. Delft University Press, Series on Mathematical Geodesy and Positioning.
- Teunissen, P.J. G. 2017. Distributional theory for the DIA method. *J Geodesy*, 91, 1-22 (doi: 10.1007/s00190-017-1045-7).
- Van Mierlo, J. 1975. Statistical analysis of geodetic networks designed for the detection of crustal movements. In: G.J. Borradaile, A.R. Ritsema, H.E. Rondeel, O.J. Simon, ed. *Progress in geodynamics*. North-Holland, Amsterdam, 52-61.
- Van Mierlo, J. 1979. Statistical analysis of geodetic measurements for the investigation of crustal movements. *Tectonophysics*, 52,457-467.
- Vaniček, P., Craymer, M.R. and Krakiwsky, E.J. 2001. Robustness analysis of geodetic horizontal networks. *J Geodesy*, 75,199-209.
- Wang, J. and Chen, Y. 1994. On the reliability measure of observations. *Acta Geodaetica et Cartographica Sinica*, 42-51.
- Wang, J. and Chen, Y. 1999. Outlier detection and reliability measures for singular adjustment models. *Geomat. Res. Aust.*, 71,57-72.
- Wang, J. and Knight, N. 2012. New outlier separability test and its application in GNSS positioning. *J Glob. Position. Syst.*, 11(1),46-57.
- Weisberg, S. and Atkinson, A.C. 1991. Simulated Annealing for the detection of Multiple Outliers using least squares and least median of squares fittin. In W. Stahel, S. Weisberg, ed. *Directions in Robust Statistics and Diagnostics*, New York, Springer-Verlag, 7-20.
- Yang, L., *et al.*, 2013. Outlier separability analysis with a multiple alternative hypotheses test. *J Geodesy*, 87(6),591-604.

Yang, L., *et al*, 2017. Extension of Internal Reliability Analysis Regarding Separability Analysis. *J. Surv. Eng*, 143(3). (doi:10.1061/(ASCE)SU.1943-5428.0000220)

Yousri, N.A., Ismail, M.A. and Kamel, M.S. 2007. Fuzzy Outlier Analysis: A Combined Clustering - Outlier Detection Approach. 2007 IEEE International Conference on Systems, Man and Cybernetics, 412-418. (doi: 10.1109/ICSMC.2007.4413873)

4.2 Avaliação do valor crítico obtido por meio do Método de Monte Carlo

Os resultados mostrados nesta e nas próximas duas seções foram frutos do artigo publicado no periódico *Remote Sensing*, intitulado “Um Método Diagnóstico de Outlier baseado em Monte Carlo para Análise de Sensibilidade”, da autoria de [Rofatto et al. \(2020b\)](#). Esse artigo encontra-se integralmente disponibilizado na Seção 4.4. Detalhes dos experimentos podem ser vistos no tópico **6. On the Probability Levels of Iterative Outlier Elimination** do artigo.

Nesta seção em específico, o objetivo foi investigar sob quais circunstâncias a dependência entre as hipóteses alternativas afeta os valores críticos associados à estatística de teste max- w do DSI.

Os valores críticos por Bonferroni são facilmente acessíveis, pois eles são obtidos da distribuição normal por meio da Equação (2.16). Entretanto, este procedimento despreza a dependência entre as hipóteses alternativas. A parte funcional e a estocástica do modelo matemático, especificamente a matriz design \mathbf{A} e a matriz covariância das observações \mathbf{Q}_e , produzem correlações entre os resíduos estimados. Para capturar essa dependência entre as hipóteses alternativas, o procedimento baseado em Monte Carlo foi fornecido no Capítulo 3, Seção 3.1.

Dois redes geodésicas, denotadas por (a) e (b), foram fornecidas para avaliar o efeito da dependência entre as hipóteses alternativas sobre valores críticos associados à estatística de teste do DSI. A rede (a) apresentou máxima correlação absoluta entre as estatísticas de teste $|\rho_{w_i, w_j}| = 0.415$, com valor médio de 0.24 e desvio-padrão de 0.14. No caso da rede (b), a máxima correlação absoluta foi $|\rho_{w_i, w_j}| = 1.00$, com valor médio absoluto de 0.72 e desvio-padrão de 0.27. Em termos qualitativos, classificamos a rede (a) como sendo de baixa correlação $|\rho_{w_i, w_j}| < 0.5$, e a rede (b) de alta correlação $|\rho_{w_i, w_j}| > 0.5$. Os valores críticos considerados foram para os seguintes níveis de significância conjunto: $\alpha' = 0.001$, $\alpha' = 0.0027$, $\alpha' = 0.01$, $\alpha' = 0.025$, $\alpha' = 0.05$ e $\alpha' = 0.1$. Os resultados mostraram que o valor crítico calculado por Monte Carlo, denotado por \hat{k} , é sempre menor ou igual ao valor crítico calculado pelo método de Bonferroni, denotado por k_{bonf} . A razão disso é que a matriz de redundância (\mathbf{R}) em (2.44) promove a correlação entre os resíduos. Com efeito, aumentam-se as chances de rejeitar erroneamente ao menos uma hipótese nula. Como a estatística de teste do DSI depende do vetor dos resíduos estimados, sempre teremos um certo grau de dependência entre as hipóteses alternativas. Portanto, com Bonferroni, o usuário não tem controle efetivo sobre o Erro Tipo I. Entretanto, existem algumas circunstâncias que Bonferroni pode ser uma boa alternativa. No caso em que temos um sistema com alta redundância ($r_i > 0.5$) e baixa correlação entre os resíduos, e para $\alpha' < 0.1$, o método de Bonferroni pode ser uma boa aproximação para controlar o

Erro Tipo I. Em geral, sob a condição da hipótese nula ser verdadeira, é menos provável que haja um $\max\text{-}w$ maior do que o previsto por Bonferroni.

4.3 Análise do efeito do valor crítico controlado sobre os níveis de probabilidade do Data-Snooping

Aqui, o objetivo foi testar se o aumento do nível de significância conjunto, ou redução do valor crítico, melhora a identificabilidade do outlier. Os valores críticos foram aqueles mesmos da seção anterior, ou seja, para os níveis de significância conjunto $\alpha' = 0.001$, $\alpha' = 0.0027$, $\alpha' = 0.01$, $\alpha' = 0.025$, $\alpha' = 0.05$ e $\alpha' = 0.1$. O procedimento descrito no Capítulo 3, Seção 3.2., Figura (10), foi aplicado para estimar os níveis de probabilidade do DSI, na presença de um outlier nos dados observados.

Os resultados mostraram que quanto menor \hat{k} , ou maior α' , maior é a correta detecção (\mathcal{P}_{CD}), e menor é o MDB. A partir daqui, mostramos os resultados em forma de símbolos matemáticos. Por exemplo, para essa relação entre \hat{k} , ou α' , e a taxa de correta detecção (\mathcal{P}_{CD}), escrevemos:

$$\uparrow \alpha' \text{ (ou } \downarrow \hat{k}) \uparrow \mathcal{P}_{CD} \downarrow MDB \quad (4.1)$$

Como sempre teremos $\hat{k} \leq k_{bonf}$, a probabilidade da correta detecção \mathcal{P}_{CD} a partir de \hat{k} é sempre maior ou no mínimo igual àquela estimada por k_{bonf} , i.e. $\mathcal{P}_{CD}(\hat{k}) \geq \mathcal{P}_{CD}(k_{bonf})$.

Em termos de correta identificação (\mathcal{P}_{CI}), os resultados revelaram que a relação $\uparrow \alpha' \text{ (ou } \downarrow \hat{k}) \uparrow \mathcal{P}_{CI} \downarrow MIB$ somente é válida para uma determinada faixa de magnitude de outlier (∇_i). Por exemplo, na Figura (15a) do artigo, quando a observação “1” (Δ_{h_1}) da rede(b), com $\max_{|\rho_{w_i, w_j}|} = 0.98$, foi contaminada por $\nabla_1 < 1.5\sigma$, \mathcal{P}_{CI} foi maior para $\alpha' = 0.1$ do que $\alpha' = 0.001$. Entretanto, houve uma inversão de \mathcal{P}_{CI} para $\nabla_1 > 1.5\sigma$, ou seja, \mathcal{P}_{CI} tornou-se maior para $\alpha' = 0.001$. A pergunta a se fazer é por que essa inversão ocorreu. Discutimos a seguir.

No caso extremo de $\nabla_1 = 1\sigma$, a diferença entre \mathcal{P}_{CD} e \mathcal{P}_{WE} para $\alpha' = 0.001$ foi menor do que $\alpha' = 0.1$, i.e.:

$$\mathcal{P}_{CD}(\alpha'=0.001) - \mathcal{P}_{WE}(\alpha'=0.001) = 20\% < \mathcal{P}_{CD}(\alpha'=0.1) - \mathcal{P}_{WE}(\alpha'=0.1) = 40\% \quad (4.2)$$

Neste caso, pode-se escrever a seguinte relação:

$$\mathcal{P}_{CD}(\alpha'=0.1) = 2 \times (\mathcal{P}_{CD}(\alpha'=0.001) - \mathcal{P}_{WE}(\alpha'=0.001)) + \mathcal{P}_{WE}(\alpha'=0.1) \quad (4.3)$$

Substituindo a Equação (4.3) na Equação (3.2) e desprezando as probabilidades pormenores, temos:

$$\begin{aligned}\mathcal{P}_{CI(\alpha'=0.1)} &= 2 \times (\mathcal{P}_{CD(\alpha'=0.001)} - \mathcal{P}_{WE(\alpha'=0.001)} + \mathcal{P}_{WE(\alpha'=0.1)} - \mathcal{P}_{WE(\alpha'=0.1)}) \\ &= 2 \times (\mathcal{P}_{CD(\alpha'=0.001)} - \mathcal{P}_{WE(\alpha'=0.001)})\end{aligned}\quad (4.4)$$

Enquanto para $\mathcal{P}_{CI(\alpha'=0.001)}$, temos:

$$\begin{aligned}\mathcal{P}_{CI(\alpha'=0.001)} &= \frac{(\mathcal{P}_{CD(\alpha'=0.1)} - \mathcal{P}_{WE(\alpha'=0.1)})}{2} + \mathcal{P}_{WE(\alpha'=0.001)} - \mathcal{P}_{WE(\alpha'=0.001)} \\ &= \frac{(\mathcal{P}_{CD(\alpha'=0.1)} - \mathcal{P}_{WE(\alpha'=0.1)})}{2}\end{aligned}\quad (4.5)$$

Logo, temos:

$$\mathcal{P}_{CI(\alpha'=0.1)} = 2 \times 20\% \approx 40\% > \mathcal{P}_{CI(\alpha'=0.001)} = 20\%/2 \approx 10\% \text{ para } \nabla_1 = 1\sigma \quad (4.6)$$

Verificamos que a relação dada acima (4.6) ocorreu para o intervalo de $\nabla_1 < 1.5\sigma$. Portanto, neste caso, tem-se: $MIB_{(\alpha'=0.1)} < MIB_{(\alpha'=0.001)}$. Portanto, $\mathcal{P}_{CI}(\hat{k}) \geq \mathcal{P}_{CI}(k_{bonf})$ para $\nabla_1 < 1.5\sigma$.

Quando $\nabla_1 > 1.5\sigma$, houve uma inversão, ou seja, $\mathcal{P}_{CI(\alpha'=0.001)} > \mathcal{P}_{CI(\alpha'=0.1)}$. Se observamos a Figura (17a) do artigo, as taxas de erro de falsa exclusão \mathcal{P}_{WE} diminuíram igualmente para todos α' quando $\nabla_1 > 1.5\sigma$. Isso significa que se escolhêssemos \hat{k} ou k_{bonf} para controlar o \mathcal{P}_{WE} a partir de $\nabla_1 > 1.5\sigma$, não haveria efeito sobre \mathcal{P}_{CI} . Ainda, se observamos a Figura (22a) do artigo, notaremos que a probabilidade da sobre-identificação com a correta identificação (\mathcal{P}_{over+}) também é nulo para $\alpha' = 0.001$. Por outro lado, no caso extremo de $\nabla_1 = 9\sigma$, por exemplo, $\mathcal{P}_{over+} = 6.7\%$.

Como $\mathcal{P}_{WE(\alpha'=0.001)} \approx \mathcal{P}_{WE(\alpha'=0.001)}$, $\mathcal{P}_{ol(\alpha'=0.001)} \approx 0$ e $\mathcal{P}_{over+(\alpha'=0.001)} \approx 0$, e considerando a relação dada em (3.2), a probabilidade da correta identificação para $\alpha' = 0.001$ foi maior do que $\alpha' = 0.1$. O aumento de α' para $\nabla_1 > 1.5\sigma$, aumentou o efeito da correlação entre as estatísticas de teste sobre os níveis de probabilidade do DSI. Consequentemente, o aumento de α' , ou seja a redução do valor crítico aumentou \mathcal{P}_{ol} e \mathcal{P}_{over+} . Podemos concluir que a inversão $\mathcal{P}_{CI(\alpha'=0.001)} > \mathcal{P}_{CI(\alpha'=0.1)}$ é devida ao aumento do efeito da correlação entre as estatísticas de teste sobre as probabilidades de \mathcal{P}_{ol} e \mathcal{P}_{over+} . Neste caso, quanto maior o valor crítico, maior será a taxa de identificação \mathcal{P}_{CI} e menor será o MIB, e vice-versa, ou seja:

$$\uparrow \alpha' \text{ (ou } \downarrow \hat{k}) \downarrow \mathcal{P}_{CI} \uparrow MIB \quad (4.7)$$

A relação acima (4.7) pode ser verificada nas Tabelas (8,9, 10 e 11) do referido artigo.

No que diz respeito à rede (a), considerando a observação 1 contaminada ∇_1 (Figura 8a) do artigo, a inversão $\mathcal{P}_{CI(\alpha'=0.001)} > \mathcal{P}_{CI(\alpha'=0.1)}$ ocorre para $\nabla_1 > 7.4\sigma$. Porém, para $\nabla_1 > 7.4\sigma$ tanto $\mathcal{P}_{CI(\alpha'=0.001)}$ quanto $\mathcal{P}_{CI(\alpha'=0.1)}$ apresentaram uma alta taxa de sucesso,

com $\mathcal{P}_{CI} = 90\%$. Para $\nabla_1 < 7.4\sigma$, $\mathcal{P}_{CI(\alpha'=0.1)}$ foi sempre superior à $\mathcal{P}_{CI(\alpha'=0.001)}$. Neste caso, portanto, a correta identificação com uso do valor crítico calculado por Monte Carlo \hat{k} será melhor do que k_{bonf} , ou seja, $\mathcal{P}_{CI}(\hat{k}) \geq \mathcal{P}_{CI}(k_{bonf})$.

4.4 Estudo da relação entre as medidas analíticas de controlabilidade e os níveis de probabilidade do Data-Snooping

Aqui, o objetivo foi estudar a associação entre as medidas analíticas de controlabilidade das observações e os níveis de probabilidade do DSI, na presença de um outlier dos dados observados. Uma das medidas de confiabilidade considerada neste estudo é o número de redundância local de cada observação (r_i). O número de redundância local indica a parcela da incerteza da observação, isto é, soma do efeitos aleatórios e do possível outlier, que é absorvida no respectivo resíduo desta observação. Os números de redundância local das observações podem ser obtidos pelos respectivos elementos da diagonal principal da matriz redundância (\mathbf{R}), dada pela seguinte expressão (KAVOURAS, 1982):

$$\mathbf{R} = \frac{1}{\sigma_0^2} \mathbf{Q}_{\hat{e}} = \mathbf{R} = \mathbf{I} - \mathbf{A}(\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \quad (4.8)$$

com $\mathbf{Q}_{\hat{e}} \in \mathbb{R}^{n \times n}$ sendo a matriz covariância dos resíduos estimados, $\mathbf{I} \in \mathbb{R}^{n \times n}$ a matriz identidade, e σ_0^2 o fator de variância a priori das observações, de caráter arbitrário, por exemplo $\sigma_0^2 = 1$.

Se as observações não forem correlacionadas (i.e., suas covariâncias são nulas), os números de redundância local (r_i) ficam restritos no seguinte intervalo:

$$0 \leq r_i \leq 1 \quad (4.9)$$

Os números de redundância local também expressam a contribuição de cada observação para a redundância global ($r = n - u$) do ajustamento, pois $r = n - u = \sum_{i=1}^n r_i$. Por exemplo se um dado $r_i = 0.5$; então cerca de 50% da incerteza total, soma da parte aleatória com o efeito de um possível outlier (ver Equação 2.44), é absorvida pelo respectivo resíduo. Portanto, quanto maior a redundância local de uma observação (r_i), mais é a controlabilidade desta observação, pois maior é a sensibilidade do seu respectivo resíduo à ocorrência de um possível outlier. Quando as observações são correlacionadas, pode-se calcular o número de confiabilidade. Para mais detalhes consultar Knight, Wang e Rizos (2010) e/ou Klein (2014).

A Figura (9) do artigo revelou que quanto maior a redundância local (r_i), menores são os indicadores de sensibilidade do DSI, ou seja:

$$\uparrow r_i \uparrow \mathcal{P}_{CD} \downarrow MDB \ \& \ \uparrow \mathcal{P}_{CI} \downarrow MIB \quad (4.10)$$

Por meio da Figura (11) e (12) do artigo foi possível obter as seguintes relações, respectivamente:

$$\uparrow \rho_{w_i, w_j} \uparrow \mathcal{P}_{WE} \text{ e } \uparrow r_i \downarrow \mathcal{P}_{WE} \quad (4.11)$$

A segunda parte da relação acima em (4.11) é válida apenas para observações com o mesmo grau de correlação entre as estatística de teste (ρ_{w_i, w_j}).

As relações dada em (4.10 e 4.11) são válidas para um sistema com alta redundância ($r_i > 0.5$) e baixa correlação entre os resíduos ($\rho_{w_i, w_j} < 0.5$). A Figura (18) do artigo revelou que a relação dada em (4.11) somente é válida a partir de uma certa magnitude de outlier (∇_i). Neste caso, observamos que, quando o efeito de todas as observações na taxa de falsa exclusão \mathcal{P}_{WE} diminui, é possível encontrar o menor outlier identificável (MIB). A razão disso é que o efeito da correlação entre os resíduos torna-se insignificante em uma certa magnitude de outlier ∇_i , o que aumenta a probabilidade da correta identificação \mathcal{P}_{CI} . Esse resultado é um dos principais destaques desta seção. Portanto, mostramos novamente a Figura (20) do artigo por meio da Figura (12). A linha tracejada preta corresponde à probabilidade da correta identificação \mathcal{P}_{CI} e do respectivo MIB para um exemplo de $\alpha' = 0.001$.

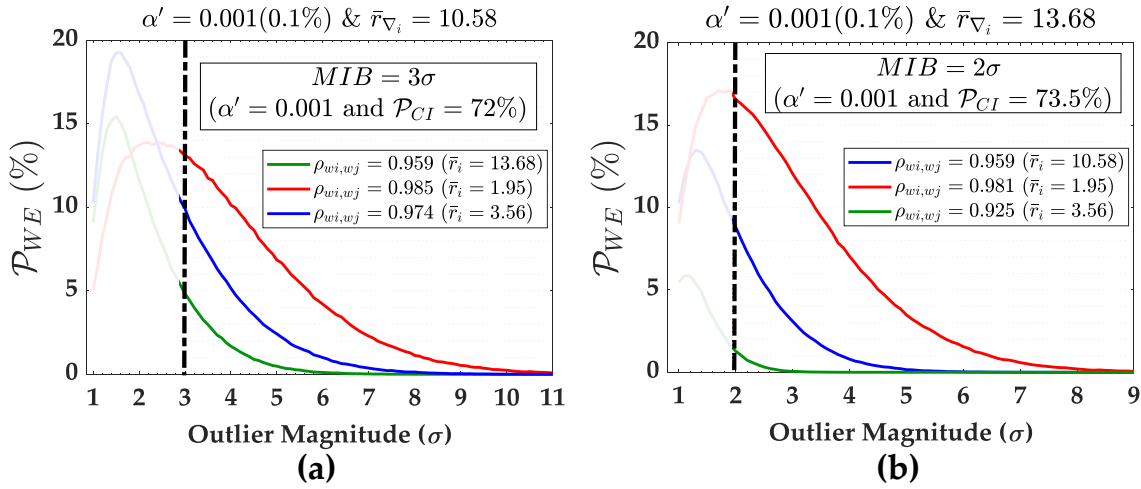


Figura 12 – Relação entre a contribuição de cada observação identificada erroneamente para a falsa exclusão (\mathcal{P}_{WE}) e a taxa de correta identificação (\mathcal{P}_{CI}).

Article

A Monte Carlo-Based Outlier Diagnosis Method for Sensitivity Analysis

Vinicius Francisco Rofatto ^{1,2,*} , Marcelo Tomio Matsuoka ^{1,2,3,4} , Ivandro Klein ^{5,6} ,
Maurício Roberto Veronez ⁴  and Luiz Gonzaga da Silveira, Jr. ⁴

¹ Graduate Program in Remote Sensing, Federal University of Rio Grande do Sul, Porto Alegre 91501970, Brazil; tomiom@ufu.br

² Institute of Geography, Federal University of Uberlandia, Monte Carmelo 38500-000, Brazil

³ Graduate Program in Agriculture and Geospatial Information, Federal University of Uberlândia, Monte Carmelo 38500-000, Brazil

⁴ Graduate Program in Applied Computing, Unisinos University, Av. Unisinos, 950, São Leopoldo 93022-000, Brazil; veronez@unisinos.br (M.R.V.); lgonzagajr@gmail.com (L.G.d.S.J.)

⁵ Department of Civil Construction, Federal Institute of Santa Catarina, Florianopolis 88020-300, Brazil; ivandroklein@gmail.com

⁶ Graduate Program in Geodetic Sciences, Federal University of Paraná, Curitiba 81531-990, Brazil

* Correspondence: vfroffatto@gmail.com or vinicius.rofatto@ufu.br

Received: 24 January 2020; Accepted: 4 March 2020; Published: 6 March 2020



Abstract: An iterative outlier elimination procedure based on hypothesis testing, commonly known as *Iterative Data Snooping (IDS)* among geodesists, is often used for the quality control of modern measurement systems in geodesy and surveying. The test statistic associated with *IDS* is the extreme normalised least-squares residual. It is well-known in the literature that critical values (quantile values) of such a test statistic cannot be derived from well-known test distributions but must be computed numerically by means of Monte Carlo. This paper provides the first results on the Monte Carlo-based critical value inserted into different scenarios of correlation between outlier statistics. From the Monte Carlo evaluation, we compute the probabilities of correct identification, missed detection, wrong exclusion, over-identifications and statistical overlap associated with *IDS* in the presence of a single outlier. On the basis of such probability levels, we obtain the Minimal Detectable Bias (MDB) and Minimal Identifiable Bias (MIB) for cases in which *IDS* is in play. The MDB and MIB are sensitivity indicators for outlier detection and identification, respectively. The results show that there are circumstances in which the larger the Type I decision error (smaller critical value), the higher the rates of outlier detection but the lower the rates of outlier identification. In such a case, the larger the Type I Error, the larger the ratio between the MIB and MDB. We also highlight that an outlier becomes identifiable when the contributions of the measures to the wrong exclusion rate decline simultaneously. In this case, we verify that the effect of the correlation between outlier statistics on the wrong exclusion rate becomes insignificant for a certain outlier magnitude, which increases the probability of identification.

Keywords: probability; hypothesis testing; outlier detection; monte carlo; quality control; control system; reliability; random number generators

1. Introduction

In recent years, Outlier Detection has been increasingly applied in sensor data processing [1–9]. Despite the countless contributions made over the years, there is continuing research on the subject, mainly because there has been an increase in computational power. One can argue that computational complexity is becoming high because of the era of information overload. However, this limitation

has been overcome over the years, mainly by the rapid development of computers, which now allow advanced computational techniques to be used efficiently on personal computers or even on handheld computers [10]. Therefore, computational complexity is no longer a bottleneck because we have fast computers and large data storage systems at our disposal [11,12].

Here, we assume that an outlier is a measurement that is so likely to be caused by a blunder that it is better to either not use it or not use it as it is [13]. Failure to identify an outlier can jeopardise the reliability level of a system. Because of its importance, outliers must be appropriately treated to ensure the quality of data analysis.

Two categories of advanced techniques for the treatment of a dataset contaminated by outliers have often been developed and applied in various situations: Robust Adjustment Procedures (see, e.g., [14–18]) and Statistical Hypothesis Testing (see, e.g., [2,12,19–23]). The first one is an estimation technique that is not unduly affected by outliers or other small departures from model assumptions. Classes of this technique include M-estimates (which follow from maximum likelihood considerations), L-Estimates (which are linear combinations of order statistics), and R-Estimates (based on statistical rank tests). Some classes of such robust adjustment methods, as well as their properties, are well known, while other methods are still being researched (see, e.g., L_1 -norm estimation [24], M-estimation [25–27], R-estimation [28–30] and those based on meta-heuristics [31]). Besides the undoubted advantages of Robust Estimation, here, we focus on the hypothesis test-based outlier. The following advantages of the outlier test were mentioned by [32]:

1. It is an opportunity to investigate the causes of outliers;
2. Identified outliers can be remeasured; and
3. If the outliers are discarded from the measurements, then standard adjustment software, which operates according to the least-squares criterion, can be used.

In this paper, we consider iterative data snooping (*IDS*), which is the most common procedure found in the geodetic practice [12,33]. Most conventional geodetic studies have a chapter on *IDS* (see, e.g., [34,35]). *IDS* has also become very popular and is routinely used in adjustment computations [36]. It is important to mention that *IDS* is not restricted to the field of geodetic statistics but is a generally applicable method [22].

IDS is an iterative outlier elimination procedure, which combines estimation, testing and adaptation [37]. Parameter estimation is often conducted in the sense of the least-squares estimation (LSE). Assuming that no outlier exists, the LSE is the best linear unbiased estimator (BLUE) [35]. The LSE has often been used in several remote sensing applications (see, e.g., [38–41]). However, outliers can inevitably occur in practice and cause the loss of the LSE BLUE-property. Then, hypothesis testing is performed with the aim of identifying any outliers that may be present in the dataset. After its identification, the suspected outlier is then excluded from the dataset as a corrective action (i.e., adaptation), and the LSE is restarted without the rejected measurement. If model redundancy permits, this procedure is repeated until no more (possible) outliers can be identified (see, e.g., [35], p. 135). Although here we restrict ourselves to the case of one outlier at a time, *IDS* can also be applied to the case of multiple (simultaneous) outliers [42]. For more details about multiple (simultaneous) outliers, refer to [43–45].

Of particular importance for quality control purposes are decision probabilities. Probability levels have already been described in the literature for the case in which data snooping is run once (i.e., only one single estimation and testing), as well as for the case in which the outlier is parameterised in the model (see, e.g., [2,19–21,23,37,46,47]). For such cases, the probability of correct detection (\mathcal{P}_{CD}) and correct identification (\mathcal{P}_{CI}) and their corresponding Minimal Detectable Bias (MDB) and Minimal Identifiable Bias (MIB) have already been described for data snooping [37,46].

The MDB is defined as the smallest value of an outlier that can be detected given a certain \mathcal{P}_{CD} . The MDB is an indicator of the sensitivity of data snooping to outlier detection and not to outlier identification. On the other hand, the MIB is defined as the smallest value of an outlier that can

be identified given a certain \mathcal{P}_{CI} ; i.e., the MIB is an indicator of the sensitivity of data snooping to outlier identification. It is important to highlight that “outlier detection” only informs us whether or not there might have been at least one outlier. However, the detection does not tell us which measurement is an outlier. The localisation of the outlier is a problem of “outlier identification”. In other words, “outlier identification” implies the execution of a search among the measurements for the most likely outlier.

However, both the MDB and MIB cannot be used as a diagnostic tool when *IDS* is in play. In this contribution, we highlight the fact that the correct outlier identification for *IDS* is not only dependent on the correct/missed detection and wrong exclusion but also other decision probabilities.

The evaluation of the probability levels associated with *IDS* is not a trivial task. When used for data snooping for a single run, the probabilities of *IDS* are multivariate integrals over complex regions [2,47]. This complexity is due to the fact that *IDS* is not only based on multiple hypothesis testing but also on multiple rounds of estimation, testing and exclusion. Because an analytical formula is not easy to compute, the Monte Carlo method should be run to obtain the probabilities and the minimal bias (MDB and MIB) indicators for *IDS*. The Monte Carlo method provides insights into these cases, in which analytical solutions are too complex to fully understand, are doubted for one reason or another or are not available [12]. The Monte Carlo method for quality control purposes has already been applied in geodesy (see, e.g., [2,10,22,23,33,46,48–51]). For in-depth coverage of Monte Carlo methods, consult, for instance, [52–54].

Recent studies by Rofatto et al. [12,55] provide an algorithm based on Monte Carlo to determine the probability levels associated with *IDS*. In that case, five classes of decisions for *IDS* are described, namely, the probability of correct identification (\mathcal{P}_{CI}), the probability of missed detection (\mathcal{P}_{MD}), the probability of wrong exclusion (\mathcal{P}_{WE}), the probability of over-identification positive (\mathcal{P}_{over+}), and the probability of over-identification negative (\mathcal{P}_{over-}), defined as follows:

- \mathcal{P}_{CI} : The probability of correctly identifying and removing an outlying measurement;
- \mathcal{P}_{MD} : The probability of not detecting the outlier (i.e., Type II decision error for *IDS*);
- \mathcal{P}_{WE} : The probability of identifying and removing a non-outlying measurement while the ‘true’ outlier remains in the dataset (i.e., Type III decision error for *IDS*);
- \mathcal{P}_{over+} : The probability of correctly identifying and removing the outlying measurement and others and
- \mathcal{P}_{over-} : The probability of identifying and removing more than one non-outlying measurement while the ‘true outlier’ remains in the dataset.

However, the procedure used by these authors [12,55] does not allow the user to control the Type I decision error (denoted by α'). The probability level α' (known as the significance level of a test) defines the size of a test and is often called the “false alarm probability”. In this paper, we highlight the fact that the test statistic associated with *IDS* does not have a known distribution, and therefore, its critical values (i.e., the percentile of its probability distribution) cannot be taken from well-known statistical tables (e.g., normal distribution).

Here, the critical value is computed by Monte Carlo such that a user-defined Type I decision error α' for *IDS* is warranted. In other words, the Type I decision error α' is effectively user-controlled when both the functional and stochastic parts of the model are taken into account. To do so, we employ the Monte Carlo method because the critical region of the test statistic associated with *IDS* is too complicated. The critical region is the subset of the measurements for which the null hypothesis \mathcal{H}_0 is rejected [12]. Therefore, the false alarm rate can be user-controlled by setting the appropriate size of the critical region.

We show that one of the advantages of having critical values based on the distribution test of *IDS* is that the dependencies between the least-squares residuals are captured by Monte Carlo simulation. In this paper, we present the first results on the Monte Carlo-based critical value in two different

scenarios of correlation between outlier test statistics. We also discuss this issue in the context of the well-known Bonferroni correction [56] to control the Type I decision error α' for *IDS*.

Moreover, herein, a new class of decision is taken into account when *IDS* is performed, which corresponds to the probability of simultaneously flagging two (or more) measurements as outliers. We call this the probability of “statistical overlap” (\mathcal{P}_{ol}). This means that \mathcal{P}_{ol} occurs in cases in which one alternative hypothesis has the same distribution as another one. In other words, these hypotheses cannot be distinguished; i.e., they are nonseparable, and an outlier cannot be identified [37].

We also investigate the probabilities of making correct decisions and the risks of incorrect decisions when *IDS* is performed in the presence of an outlier in two different scenarios of correlation between outlier test statistics. On the basis of the probability levels associated with *IDS* (i.e., \mathcal{P}_{CI} , $\mathcal{P}_{MD}/\mathcal{P}_{CD}$, \mathcal{P}_{WE} , \mathcal{P}_{over+} , \mathcal{P}_{over-} and \mathcal{P}_{ol}), we also show how to find the two sensitivity indicators MDB and MIB for *IDS*. We also analyse the relationship between the sensitivity indicators MDB and MIB for *IDS*.

2. Binary Hypothesis Testing versus Multiple Hypothesis Testing: True Data Snooping

Random measurement errors in a system are unavoidable. The stochastic properties of measurement errors are directly associated with the assumption of the probability distribution of these errors. In geodesy and many other scientific branches, the well-known normal distribution is one of the most used measurement error models. Its choice is further justified by both the central limit theorem and the maximum entropy principle. Some alternative measurement error models can be found in [11].

Therefore, the null hypothesis, denoted by \mathcal{H}_0 , is formulated under the condition that random errors are normally distributed with expectation zero. In other words, the model associated with the null hypothesis \mathcal{H}_0 consists of the one believed to be valid under normal working conditions, i.e., in the absence of outliers. When it is assumed to be ‘true’, this model is used to estimate unknown parameters, usually in a least-squares approach. Thus, the null hypothesis \mathcal{H}_0 of the standard Gauss–Markov model in the linear or linearised form is given by [34]

$$\mathcal{H}_0 : \mathbb{E}\{\mathbf{y}\} = \mathbf{A}\mathbf{x} + \mathbb{E}\{\mathbf{e}\} = \mathbf{A}\mathbf{x}; \mathbb{D}\{\mathbf{y}\} = \mathbf{Q}_e \quad (1)$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator, $\mathbb{D}\{\cdot\}$ is the dispersion operator, $\mathbf{y} \in \mathbb{R}^{n \times 1}$ is the vector of measurements, $\mathbf{A} \in \mathbb{R}^{n \times u}$ is the Jacobian matrix (also called the design matrix) of full rank u , $\mathbf{x} \in \mathbb{R}^{u \times 1}$ is the unknown parameter vector, $\mathbf{e} \in \mathbb{R}^{n \times 1}$ is the unknown vector of measurement errors and $\mathbf{Q}_e \in \mathbb{R}^{n \times n}$ is the positive-definite covariance matrix of the measurements \mathbf{y} .

Under normal working conditions (i.e., \mathcal{H}_0), the measurement error model is then given by

$$\mathbf{e} \sim N(\mathbf{0}, \mathbf{Q}_e), \quad (2)$$

Here, we confine ourselves to the case in which \mathbf{A} and \mathbf{Q}_e have full column rank.

The best linear unbiased estimator (BLUE) of \mathbf{e} under \mathcal{H}_0 is the well-known estimated least-squares residual vector $\hat{\mathbf{e}} \in \mathbb{R}^{n \times 1}$, which is given by

$$\begin{aligned} \hat{\mathbf{e}} &= \mathbf{y} - \mathbf{A}\hat{\mathbf{x}} \\ &= \mathbf{y} - \mathbf{A}(\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{W} \mathbf{y}) \\ &= \mathbf{A}\mathbf{x} + \mathbf{e} - \mathbf{A}(\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{W} (\mathbf{A}\mathbf{x} + \mathbf{e})) \\ &= \mathbf{e} - \mathbf{A}(\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{W} \mathbf{e}) \\ &= (\mathbf{I} - \mathbf{A}(\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W}) \mathbf{e} \\ &= \mathbf{R} \mathbf{e}, \end{aligned} \quad (3)$$

with $\hat{\mathbf{x}} \in \mathbb{R}^{u \times 1}$ being the BLUE of \mathbf{x} under \mathcal{H}_0 ; $\mathbf{W} \in \mathbb{R}^{n \times n}$ is the known matrix of weights, taken as $\mathbf{W} = \sigma_0^2 \mathbf{Q}_e^{-1}$, where σ_0^2 is the variance factor, $\mathbf{I} \in \mathbb{R}^{n \times n}$ is the identity matrix and $\mathbf{R} \in \mathbb{R}^{n \times n}$ is known

as the redundancy matrix. The \mathbf{R} matrix is an orthogonal projector that projects onto the orthogonal complement of the range space of \mathbf{A} .

We restrict ourselves to regular models, and therefore, the degrees of freedom r (redundancy) of the model under \mathcal{H}_0 (Equation (1)) is

$$r = \text{rank}(\mathbf{Q}_{\hat{e}}) = n - \text{rank}(\mathbf{A}) = n - u, \text{ where} \quad (4)$$

$$\mathbf{Q}_{\hat{e}} = \mathbf{Q}_e - \sigma_0^2 \mathbf{A}(\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \quad (5)$$

On the other hand, an alternative model is proposed when there are doubts about the reliability level of the model under \mathcal{H}_0 . Here, we assume that the validity of the null hypothesis \mathcal{H}_0 in Equation (1) can be violated if the dataset is contaminated by outliers. The model in an alternative hypothesis, denoted by \mathcal{H}_A , is to oppose Equation (1) by an extended model that includes the unknown vector $\nabla \in \mathbb{R}^{q \times 1}$ of deterministic bias parameters as follows ([20,35]):

$$\mathcal{H}_A : \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{C}\nabla + \mathbf{e} = \begin{pmatrix} \mathbf{A} & \mathbf{C} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \nabla \end{pmatrix} + \mathbf{e}, \quad (6)$$

where $\mathbf{C} \in \mathbb{R}^{n \times q}$ is the matrix that relates bias parameters, i.e., the values of the outliers to observations. We restrict ourselves to the matrix $\begin{pmatrix} \mathbf{A} & \mathbf{C} \end{pmatrix}$ having full column rank, such that

$$r = \text{rank} \begin{pmatrix} \mathbf{A} & \mathbf{C} \end{pmatrix} = u + q \leq n \quad (7)$$

One of the most used procedures based on hypothesis testing for outliers in linear (or linearised) models is the well-known data snooping method [19,20]. This procedure consists of screening each individual measurement for the presence of an outlier [42]. In that case, data snooping is based on a local model test, such that $q = 1$, and therefore, the n alternative hypothesis is expressed as

$$\mathcal{H}_A^{(i)} : \mathbf{y} = \mathbf{A}\mathbf{x} + c_i \nabla_i + \mathbf{e} = \begin{pmatrix} \mathbf{A} & c_i \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \nabla_i \end{pmatrix} + \mathbf{e}, \forall i = 1, \dots, n \quad (8)$$

Now, matrix \mathbf{C} in Equation (6) is reduced to a canonical unit vector c_i , which consists exclusively of elements with values of 0 and 1, where 1 means that the i th bias parameter of magnitude ∇_i affects the i th measurement, and 0 means otherwise. In that case, the rank of $\begin{pmatrix} \mathbf{A} & c_i \end{pmatrix} \in \mathbb{R}^{n \times (u+1)}$ and the vector ∇ in Equation (6) reduces to a scalar ∇_i in Equation (8), i.e., $c_i = \begin{pmatrix} 0 & 0 & 0 & \dots & 1^{i\text{th}} & 0 & \dots & 0 \end{pmatrix}^T$. When $q = n - u$, an overall model test is performed. For more details about the overall model test, see, for example, [46,47].

Note that the alternative hypothesis $\mathcal{H}_A^{(i)}$ in Equation (8) is formulated under the condition that the outlier acts as a systematic effect by shifting the random error distribution under \mathcal{H}_0 by its own value [13]. In other words, the presence of an outlier in a dataset can cause a shift of the expectation under \mathcal{H}_0 to a nonzero value. Therefore, hypothesis testing is often employed to check whether the possible shifting of the random error distribution under \mathcal{H}_0 by an outlier is, in fact, a systematic effect (bias) or merely a random effect. This hypothesis test-based approach is called the *mean-shift model* [20]. The mean-shift model has been widely employed in a variety of applications, such as structural deformation analyses, sensor data processing, the integrity monitoring of GNSS (Global Navigation Satellite System) models and the design and quality control of geodetic networks (see, e.g., [1,3,6,8,12,19–22,45,51,57–61]). The alternative to the mean-shift model is variance inflation. Until now, it has been rarely used in geodesy because it is more difficult to derive a powerful test and a reliability theory for it [12,13,62].

2.1. Binary Hypothesis Testing

In the context of the mean-shift model, the test statistic involved in data snooping is given by the normalised least-squares residual, denoted by w_i . This test statistic, also known as Baarda’s w -test, is given as follows:

$$w_i = \frac{c_i^T Q_e^{-1} \hat{e}}{\sqrt{c_i^T Q_e^{-1} Q_{\hat{e}} Q_e^{-1} c_i}}, \forall i = 1, \dots, n \tag{9}$$

Then, a test decision is performed as [63]

$$\text{Accept } \mathcal{H}_0 \text{ if } |w_i| \leq k, \text{ reject otherwise in favour of } \mathcal{H}_A^{(i)} \tag{10}$$

Note that the decision rule (10) says that if the Baarda’s w -test statistic is larger than some critical value k , i.e., a percentile of its probability distribution, then we reject the null hypothesis in favour of the alternative hypothesis. This is a special case of testing the null hypothesis \mathcal{H}_0 against only one single alternative hypothesis $\mathcal{H}_A^{(i)}$, and therefore, the rejection of the null hypothesis automatically implies the acceptance of the alternative hypothesis and vice versa [46,47]. In other words, the outlier detection automatically implies outlier identification and vice versa. This is because the formulation of the alternative hypothesis $\mathcal{H}_A^{(i)}$ is based on the condition that an outlier exists and is located at a pre-specified position in the dataset. In other words, the alternative hypothesis in a binary test says that “a specific measurement is an outlier”.

Because Baarda’s w -test in its essence is based on binary hypothesis testing, in which one decides between the null hypothesis \mathcal{H}_0 and only one single alternative hypothesis $\mathcal{H}_A^{(i)}$ of (8), it may lead to wrong decisions of Type I and Type II. The probability of a Type I Error α_0 is the probability of rejecting the null hypothesis \mathcal{H}_0 when it is true, whereas the probability of a Type II error β_0 is the probability of failing to reject the null hypothesis \mathcal{H}_0 when it is false (note: the index ‘0’ represents the case in which a single hypothesis is tested). Instead of α_0 and β_0 , there is the confidence level $CL = 1 - \alpha_0$ and the power of the test $\gamma_0 = 1 - \beta_0$, respectively. The first deals with the probability of accepting a true null hypothesis \mathcal{H}_0 ; the second addresses the probability of correctly accepting the alternative hypothesis $\mathcal{H}_A^{(i)}$. In that case, given a probability of a Type I decision error α_0 , we find the critical value k_0 as follows:

$$k_0 = \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \tag{11}$$

where Φ^{-1} denotes the inverse of the cumulative distribution function (cdf) of the two-tailed standard normal distribution $N(0, 1)$.

The normalised least-squares residual w_i follows a standard normal distribution with the expectation that $\mathbb{E}\{w_i\} = 0$ if \mathcal{H}_0 holds true (there is no outlier). On the other hand, if the system is contaminated with a single outlier at the i th location of the dataset (i.e., under $\mathcal{H}_A^{(i)}$), then the expectation of w_i is

$$\mathbb{E}\{w_i\} = \sqrt{\lambda_0} = \sqrt{c_i^T Q_e^{-1} Q_{\hat{e}} Q_e^{-1} c_i \nabla_i^2} \tag{12}$$

where λ_0 is the non-centrality parameter for $q = 1$. Note, therefore, that there is an outlier that causes the expectation of w_i to become $\sqrt{\lambda_0}$. The square-root of the non-centrality parameter $\sqrt{\lambda_0}$ in Equation (12) represents the expected mean shift of a specific w -test. In such a case, the term $c_i^T Q_e^{-1} Q_{\hat{e}} Q_e^{-1} c_i$ in Equation (12) is a scalar, and therefore, it can be rewritten as follows [64]:

$$|\nabla_i| = MDB_{0(i)} = \sqrt{\frac{\lambda_0}{c_i^T Q_e^{-1} Q_{\hat{e}} Q_e^{-1} c_i}}, \forall i = 1, \dots, n \tag{13}$$

where $|\nabla_i|$ is the Minimal Detectable Bias ($MDB_{0(i)}$) for the case in which there is only one single alternative hypothesis, which can be computed for each of the n alternative hypotheses according to Equation (8).

For a single outlier, the variance of an estimated outlier, denoted by $\sigma_{\nabla_i}^2$, is

$$\sigma_{\nabla_i}^2 = \left(\mathbf{c}_i^T \mathbf{Q}_e^{-1} \mathbf{Q}_e \mathbf{Q}_e^{-1} \mathbf{c}_i \right)^{-1}, \forall i = 1, \dots, n \quad (14)$$

Thus, the MDB can also be written as

$$MDB_{0(i)} = \sigma_{\nabla_i} \sqrt{\lambda_0}, \forall i = 1, \dots, n \quad (15)$$

where $\sigma_{\nabla_i} = \sqrt{\sigma_{\nabla_i}^2}$ is the standard deviation of estimated outlier ∇_i .

The MDB in Equations (13) or (15) of an alternative hypothesis is the smallest-magnitude outlier that can lead to the rejection of the null hypothesis \mathcal{H}_0 for a given α_0 and β_0 . Thus, for each model of the alternative hypothesis $\mathcal{H}_A^{(i)}$, the corresponding MDB can be computed [12,49,65]. The limitation of this MDB is that it was initially developed for the binary hypothesis testing case. In that case, the MDB is a sensitivity indicator of Baarda's w -test when only one single alternative hypothesis is taken into account. In this article, we are confined to multiple alternative hypotheses. Therefore, both the MDB and MIB are computed by considering the case of multiple hypothesis testing.

2.2. Multiple Hypothesis Testing

The alternative hypothesis in Equation (8) has been formulated under the assumption that the measure y_i for some fixed i is an outlier. From a practical point of view, however, we do not know which measurement is an outlier. Therefore, a more appropriate alternative hypothesis would be [22] "There is at least one outlier in the vector of measurements y_i ". Now, we are interested in knowing which of the alternative hypotheses may lead to the rejection of the null hypothesis with a certain probability. This means testing \mathcal{H}_0 against $\mathcal{H}_A^{(1)}, \mathcal{H}_A^{(2)}, \mathcal{H}_A^{(3)}, \dots, \mathcal{H}_A^{(n)}$. This is known as multiple hypothesis testing (see, e.g., [1,2,12,21,23,37,46,66–69]). In that case, the test statistic coming into effect is the maximum absolute Baarda's w -test value (denoted by $\max-w$), which is computed as [12]

$$\max-w = \max_{i \in \{1, \dots, n\}} |w_i| \quad (16)$$

The decision rule for this case is given by

$$\begin{aligned} & \text{Accept } \mathcal{H}_0 \text{ if } \max-w \leq \hat{k} \\ & \text{Otherwise,} \\ & \text{Accept } \mathcal{H}_A^{(i)} \text{ if } \max-w > \hat{k} \end{aligned} \quad (17)$$

The decision rule in 17 says that if none of the n w -tests get rejected, then we accept the null hypothesis \mathcal{H}_0 . If the null hypothesis \mathcal{H}_0 is rejected in any of the n tests, then one can only assume that detection occurred. In other words, if the $\max-w$ is larger than some percentile of its probability distribution (i.e., some critical value \hat{k}), then there is evidence that there is an outlier in the dataset. Therefore, "outlier detection" only informs us whether the null hypothesis \mathcal{H}_0 is accepted or not.

However, the detection does not tell us which alternative hypothesis $\mathcal{H}_A^{(i)}$ would have led to the rejection of the null hypothesis \mathcal{H}_0 . The localisation of the alternative hypothesis, which would have rejected the null hypothesis, is a problem of "outlier identification". Outlier identification implies the execution of a search among the measurements for the most likely outlier. In other words, one seeks to find which of Baarda's w -test is the maximum absolute value $\max-w$ and if that $\max-w$ is greater than some critical value \hat{k} .

Therefore, the data snooping procedure of screening measurements for possible outliers is actually an important case of multiple hypothesis testing and not single hypothesis testing. Moreover, note that outlier identification only happens when outlier detection necessarily exists; i.e., “outlier identification” only occurs when the null hypothesis \mathcal{H}_0 is rejected. However, correct detection does not necessarily imply correct identification [2,12,46].

3. Probability Levels of Data Snooping for a Single Run under Multiple Alternative Hypotheses

Two sides of the multiple testing problem can be formulated: one under the reality of the null hypothesis \mathcal{H}_0 , i.e., the event that there is no outlier in the dataset, and another one coinciding with the alternative hypothesis $\mathcal{H}_A^{(i)}$, i.e., the event that there is an outlier. The probability levels associated with data snooping for both events are presented in Table 1.

Table 1. Probability levels associated with data snooping under multiple alternative hypotheses.

Reality Unknown	Result of the Test				
	\mathcal{H}_0	$\mathcal{H}_A^{(1)}$	$\mathcal{H}_A^{(2)}$...	$\mathcal{H}_A^{(n)}$
\mathcal{H}_0	Correct decision $1-\alpha'$	Type I Error α_{01}	Type I Error α_{02}	...	Type I Error α_{0n}
$\mathcal{H}_A^{(1)}$	Type II error β_{10}	Correct identification $1-\beta_{11}$	Type III error κ_{12}	...	Type III error κ_{1n}
$\mathcal{H}_A^{(2)}$	Type II error β_{20}	Type III error κ_{21}	Correct identification $1-\beta_{22}$...	Type III error κ_{2n}
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
$\mathcal{H}_A^{(n)}$	Type II error β_{n0}	Type III error κ_{n1}	Type III error κ_{n2}	...	Correct identification $1-\beta_{nn}$

3.1. On the Scenario Coinciding with the Null Hypothesis \mathcal{H}_0

For the scenario coinciding with the null hypothesis \mathcal{H}_0 , there is the probability of incorrectly identifying at least one alternative hypothesis. This type of wrong decision is known as the *family-wise error rate (FWE)*. The *FWE* is defined as

$$FWE = \alpha_{0i} = \mathcal{P} \left(|w_i| > |w_j| \forall j, |w_i| > k(i \neq j) \mid \mathcal{H}_0 : true \right), \forall i = 1, \dots, n \tag{18}$$

The probability of accepting the null hypothesis in test i is $1-\alpha, \forall i = 1, \dots, n$, where α is the significance level or size of the test for single hypothesis testing. The classical and well-known procedure to control the *FWE* is the Bonferroni correction [56]. If all tests are mutually independent, then the probability that a true \mathcal{H}_0 is accepted in each test is approximately

$$(1 - \alpha)^n = 1 - \alpha' \tag{19}$$

where α' is the Type I Error for the entire dataset. Thus, we have

$$\alpha = 1 - (1 - \alpha')^{1/n} \tag{20}$$

which is approximately

$$\alpha = \frac{\alpha'}{n} \tag{21}$$

The quantity in Equation (21) is just equal to the upper bound of the Bonferroni inequality, i.e., $\alpha' \leq n\alpha$ [56].

Controlling the *FWE* at a pre-specified level α' corresponds to controlling the probability of a Type I decision error when carrying out a single test. In other words, one uses a global Type I Error rate α'

that combines all tests under consideration instead of an individual error rate α that only considers one test at a time [69]. In that case, the critical value k_{bonf} is computed as

$$k_{bonf} = \Phi^{-1} \left(1 - \frac{\alpha'}{2n} \right) \quad (22)$$

For single hypothesis testing, given a probability of a Type I decision error α_0 , it is easier for us to find the critical value using Equation (11). On the other hand, the rate of Type I decision errors for multiple testing, α' , cannot be directly controlled by the user. One can argue about the application of Bonferroni [56] using Equation (22). However, Bonferroni is a good approximation for the case in which alternative hypotheses are independent. In practice, however, the test results always depend on each other to some degree because we always have a correlation between w -tests. The correlation coefficient between any Baarda's w -test statistic (denoted by ρ_{w_i, w_j}), such as w_i and w_j , is given by [21]

$$\rho_{w_i, w_j} = \frac{c_i^T Q_e^{-1} Q_\delta Q_e^{-1} c_j}{\sqrt{c_i^T Q_e^{-1} Q_\delta Q_e^{-1} c_i} \sqrt{c_j^T Q_e^{-1} Q_\delta Q_e^{-1} c_j}}, \forall (i \neq j) \quad (23)$$

The correlation coefficient ρ_{w_i, w_j} can assume values within the range $[-1, 1]$.

Here, the extreme normalised residuals max- w (i.e., maximum absolute) in Equation (16) are treated directly as a test statistic. Note that when using Equation (16) as a test statistic, the decision rule is based on a one-sided test of the form $\max-w \leq \hat{k}$. However, the distributions of max- w cannot be derived from well-known test distributions (e.g., normal distribution). Therefore, critical values cannot be taken from a statistical table but must be computed numerically. This problem has already been addressed by Lehmann [22]. In that case, the dependencies between residuals are not neglected because the critical values are based on the distribution of max- w , which depends on the correlation between w -test statistics ρ_{w_i, w_j} .

According to Equation (23), the correlation ρ_{w_i, w_j} depends on the matrices A and Q_e , and therefore, the distribution of max- w also depends on these matrices. In other words, the critical value depends on the uncertainty of the measurement sensor and the mathematical model of the problem.

In order to guarantee the user-defined Type I decision error α' for data snooping, the critical value must be computed by Monte Carlo.

The key of Monte Carlo is artificial random numbers (ARN) [70], which are called 'artificial' because the random numbers are generated using a deterministic process. A random number generator is a technology designed to generate a deterministic sequence of numbers that do not have any pattern and therefore appear to be random. It is 'random' in the sense that the sequence of numbers generated passes statistical tests for randomness. For this reason, random number generators are typically referred to as pseudo-random number generators (PRNGs).

A PRNG simulates a sequence of independent and identically distributed (i.i.d.) numbers chosen uniformly between 0 and 1. PRNGs are part of many machine learning and data mining techniques. In a simulation, a PRNG is implemented as a computer algorithm in some programming language and is made available to the user via procedure calls or icons [71]. A good generator produces numbers that are not distinguishable from truly random numbers in limited computation time. This is particularly true for Mersenne Twister, a popular generator with the long period length of $2^{199371} - 1$ [72].

In essence, Monte Carlo replaces random variables with computer ARN, probabilities with relative frequencies and expectations with arithmetic means over large sets of such numbers [12]. A computation with one set of ARN is a Monte Carlo experiment [33].

The procedure to compute the critical value of max- w is given step-by-step as follows:

1. Specify the probability density function (pdf) of the w -test statistics. The pdf assigned to the w -test statistics under an \mathcal{H}_0 -distribution is

$$(w_1, w_2, w_3, \dots, w_n)^T \sim N(\mathbf{0}, \mathcal{R}_w) \tag{24}$$

where $\mathcal{R}_w \in \mathbb{R}^{n \times n}$ is the correlation matrix with the main diagonal elements equal to 1, and the off-diagonal elements are the correlation between the w -test statistics computed by Equation (23).

2. In order to have w -test statistics under \mathcal{H}_0 , uniformly distributed random number sequences are produced by the Mersenne Twister algorithm, and then they are transformed into a normal distribution by using the Box–Muller transformation [73]. Box–Muller has already been used in geodesy for Monte Carlo experiments [22,33,74]. Therefore, a sequence of m random vectors from the pdf assigned to the w -test statistics is generated according to Equation (24). In that case, we have a sequence of m vectors of the w -test statistics as follows:

$$\left[(w_1, w_2, w_3, \dots, w_n)^{T(1)}, (w_1, w_2, w_3, \dots, w_n)^{T(2)}, \dots, (w_1, w_2, w_3, \dots, w_n)^{T(m)} \right] \tag{25}$$

3. Compute the test statistic by Equation (16) for each sequence of w -test statistics. Thus, we have

$$\left(\max_{i \in \{1, \dots, n\}} |w_i|^{(1)}, \max_{i \in \{1, \dots, n\}} |w_i|^{(2)}, \dots, \max_{i \in \{1, \dots, n\}} |w_i|^{(m)} \right) \tag{26}$$

4. Sort in ascending order the maximum test statistic in Equation (26), getting a sorted vector \tilde{w} , such that

$$\tilde{w}^{(1)} < \tilde{w}^{(2)}, \tilde{w}^{(3)}, \dots, < \tilde{w}^{(m)} \tag{27}$$

The sorted values \tilde{w} in Equation (27) provide a discrete representation of the cumulative density function (cdf) of the maximum test statistic $\max\text{-}w$.

5. Determine the critical value \hat{k} as follows:

$$\hat{k} = \tilde{w}_{\lfloor (1-\alpha') \times m \rfloor} \tag{28}$$

where $\lfloor \cdot \rfloor$ denotes rounding down to the next integer that indicates the position of the selected elements in the ascending order of \tilde{w} . This position corresponds to a critical value for a stipulated overall false alarm probability α' . This can be done for a sequence of values α' in parallel.

It is important to mention that the probability of a Type I decision error for multiple testing α' is larger than that of Type I for single testing α_0 . This is because the critical region in multiple testing is larger than that in single hypothesis testing.

3.2. On the Scenario Coinciding with the Alternative Hypothesis $\mathcal{H}_A^{(i)}$

The other side of the multiple testing problem is the situation in which there is an outlier in the dataset. In that case, apart from Type I and Type II errors, there is a third type of wrong decision associated with Baarda’s w -test. Baarda’s w -test can also flag a non-outlying observation while the ‘true’ outlier remains in the dataset. We are referring to the Type III error [67], also referred to as the probability of wrong identification (\mathcal{P}_{WI}). The description of the Type III error (denoted by κ_{ij} in Table 1) involves a separability analysis between alternative hypotheses [2,21,23,66]. Therefore, we are now interested in the identification of the correct alternative hypothesis. In that case, the non-centrality parameter in Equation (12) is not only related to the sizes of Type I and Type II decision errors but also dependent on the correlation coefficient ρ_{w_i, w_j} given by Equation (23).

On the basis of the assumption that one outlier is in the i th position of the dataset (i.e., $\mathcal{H}_A^{(i)}$ is ‘true’), the probability of a Type II error (also referenced as the probability of “missed detection”, denoted by \mathcal{P}_{MD}) for multiple testing is

$$\mathcal{P}_{MD} = \beta_{i0} = \mathcal{P} \left(\bigcap_{i=1}^n |w_i| \leq \hat{k} \mid \mathcal{H}_A^{(i)} : true \right), \tag{29}$$

and the size of a Type III wrong decision (also called “misidentification”, denoted by \mathcal{P}_{WI}) is given by

$$\mathcal{P}_{WI} = \sum_{i=1}^n \kappa_{ij} = \sum_{i=1}^n \mathcal{P} \left(|w_j| > |w_i| \forall i, |w_j| > \hat{k} (i \neq j) \mid \mathcal{H}_A^{(i)} : true \right) \tag{30}$$

On the other hand, the probability of correct identification (denoted by \mathcal{P}_{CI}) is

$$\mathcal{P}_{CI} = 1 - \beta_{ii} = \mathcal{P} \left(|w_i| > |w_j| \forall j, |w_i| > \hat{k} (i \neq j) \mid \mathcal{H}_A^{(i)} : true \right) \tag{31}$$

with

$$1 - \mathcal{P}_{CI} = \beta_{ii} = \beta_{i0} + \sum_{i=1}^n \kappa_{ij}, \text{ for } (i \neq j) \tag{32}$$

Note that the three probabilities of missed detection \mathcal{P}_{MD} , wrong identification \mathcal{P}_{WI} and correct identification \mathcal{P}_{CI} sum up to unity: i.e., $\mathcal{P}_{MD} + \mathcal{P}_{WI} + \mathcal{P}_{CI} = 1$.

The probability of correct detection \mathcal{P}_{CD} is the sum of the probability of correct identification \mathcal{P}_{CI} (selecting a correct alternative hypothesis) and the probability of misidentification \mathcal{P}_{WI} (selecting one of the $n - 1$ other hypotheses), i.e.,

$$\mathcal{P}_{CD} = \mathcal{P}_{CI} + \mathcal{P}_{WI} \tag{33}$$

The probability of wrong identification \mathcal{P}_{WI} is identically zero, $\mathcal{P}_{WI} = 0$, when the correlation coefficient is exactly zero, $\rho_{w_i, w_j} = 0$. In that case, we have

$$\mathcal{P}_{CD} = \mathcal{P}_{CI} = 1 - \mathcal{P}_{MD} \tag{34}$$

The relationship given in Equation (34) would only happen if one neglected the nature of the dependence between alternative hypotheses. In other words, this relationship is valid for the special case of testing the null hypothesis \mathcal{H}_0 against only one single alternative hypothesis $\mathcal{H}_A^{(i)}$.

Since the critical region in multiple hypothesis testing is larger than that in single hypothesis testing, the Type II decision error (i.e., \mathcal{P}_{MD}) for the multiple test becomes smaller [12]. This means that the correct detection in binary hypothesis testing (γ_0) is smaller than the correct detection \mathcal{P}_{CD} under multiple hypothesis testing, i.e.,

$$\mathcal{P}_{CD} > \gamma_0 \tag{35}$$

Detection is easier in the case of multiple hypothesis testing than single hypothesis testing. However, the probability of correct detection \mathcal{P}_{CD} under multiple testing is spread out over all alternative hypotheses, and therefore, identifying is harder than detecting. From Equation (33), it is also noted that detection does not depend on identification. However, outlier identification depends on correct outlier detection. Therefore, we have the following inequality:

$$\mathcal{P}_{CI} \leq \mathcal{P}_{CD} \tag{36}$$

Note that the probability of correct identification \mathcal{P}_{CI} depends on the probability of missed detection \mathcal{P}_{MD} and wrong identification \mathcal{P}_{WI} for the case in which data snooping is run only once, i.e., a single round of estimation and testing. However, in this paper, we deal with data snooping in its iterative form (i.e., *IDS*), and therefore, the probability of correct identification \mathcal{P}_{CI} depends on other decision rules.

4. On the Probability Levels of Iterative Outlier Elimination and Its Sensitivity Indicators

In the previous section, probability levels are described for the case in which the data snooping procedure is applied only once according to the detector given by Equation (16). In practice, however, data snooping is applied iteratively: after identification and elimination of a single outlier, the model is reprocessed, and outlier identification is restarted. This procedure of iterative outlier elimination is known as iterative data snooping (IDS) [35]. Therefore, IDS is not only a case of multiple hypothesis testing but also a case of multiple runs of estimation, testing and adaptation. In that case, adaptation consists of removing a possible outlier.

Rofatto et al. [12,55] showed how to compute the probability levels associated with the IDS procedure. They introduced two new classes of wrong decisions for IDS, namely, over-identification positive and over-identification negative. The first is the probability of IDS flagging the outlier and good observations simultaneously. The second is the probability of IDS flagging only the good observations as outliers (more than one) while the outlier remains in the dataset.

This paper extends the current decision errors of IDS for the case in which there is a single outlier in the dataset. In addition to the probability levels described so far, there is the probability that the detector in (16) simultaneously flags two (or more) observations during a round of IDS. Here, this is referred to as *statistical overlap*. Statistical overlap occurs when two (or more) Baarda's w -test statistics are equal. For instance, if the correlation coefficient between two w -test statistics (ρ_{w_i, w_j}) were exactly 1.00 (or -1.00), i.e., if $\rho_{w_i, w_j} = \pm 1.00$, then the alternative hypothesis, say, $\mathcal{H}_A^{(i)}$, would have the same distribution as another one, $\mathcal{H}_A^{(j)}$. This would mean that those hypotheses would not be distinguished, i.e., they would not be separable, and an outlier would not be identified [2]. Note that the correlation ρ_{w_i, w_j} provides an indication of whether or not the system redundancy is sufficient to identify an outlier. When the correlation coefficient between two w -test statistics is exactly 1.00 (or -1.00), i.e., $\rho_{w_i, w_j} = \pm 1.00$, a statistical overlap \mathcal{P}_{ol} is expected to occur. We further discuss \mathcal{P}_{ol} when we present the results.

In contrast to the data snooping single run, the success rate of correct detection \mathcal{P}_{CD} for IDS depends on the sum of the probabilities of correct identification (\mathcal{P}_{CI}), wrong exclusion (\mathcal{P}_{WE}), over-identification cases (\mathcal{P}_{over+} and \mathcal{P}_{over-}), and statistical overlap (\mathcal{P}_{ol}), i.e.,

$$\mathcal{P}_{CD} = 1 - \mathcal{P}_{MD} = \mathcal{P}_{CI} + \mathcal{P}_{WE} + \mathcal{P}_{over+} + \mathcal{P}_{over-} + \mathcal{P}_{ol} \quad (37)$$

It is important to mention that the probability of correct detection is the complement of the probability of missed detection. Note from Equation (39) that the probability of correct detection \mathcal{P}_{CD} is available even for cases in which the identification rate is null, $\mathcal{P}_{CI} = 0$. However, the probability of correct identification (\mathcal{P}_{CI}) necessarily requires that the probability of correct detection \mathcal{P}_{CD} be greater than zero. For the same reasons given for the data snooping single run in the previous section, detecting is easier than identifying. In that case, we have the following relationship for the success rate of correct outlier identification \mathcal{P}_{CI} :

$$\mathcal{P}_{CI} = \mathcal{P}_{CD} - (\mathcal{P}_{WE} + \mathcal{P}_{over+} + \mathcal{P}_{over-} + \mathcal{P}_{ol}), \quad (38)$$

such as

$$\exists(\mathcal{P}_{CI}) \in [0, 1] \iff (\mathcal{P}_{CD}) > 0 \quad (39)$$

It is important to mention that the wrong exclusion \mathcal{P}_{WE} describes the probability of identifying and removing a non-outlying measurement while the 'true' outlier remains in the dataset. In other words, \mathcal{P}_{WE} is the Type III decision error for IDS). The overall wrong exclusion \mathcal{P}_{WE} is the result of the sum of each individual contribution to \mathcal{P}_{WE} , i.e.,

$$\mathcal{P}_{WE} = \sum_{i=1}^{n-1} \mathcal{P}_{WE(i)} \quad (40)$$

We can also compute a weighting factor, denoted by $p_{i(\mathcal{P}_{WE})}$, for each individual contribution to \mathcal{P}_{WE} as follows:

$$p_{i(\mathcal{P}_{WE})} = \frac{\mathcal{P}_{WE(i)}}{\mathcal{P}_{WE}}, \forall i = 1, \dots, n - 1, \tag{41}$$

so that

$$\sum_{i=1}^{n-1} p_{i(\mathcal{P}_{WE})} = \frac{\sum_{i=1}^{n-1} \mathcal{P}_{WE(i)}}{\mathcal{P}_{WE}} \tag{42}$$

The weighting factor $p_{i(\mathcal{P}_{WE})}$ is within a range of [0,1].

On the basis of the probability levels of correct detection \mathcal{P}_{CD} and correct identification \mathcal{P}_{CI} , the sensitivity indicators of minimal biases—Minimal Detectable Bias (MDB) and Minimal Identifiable Bias (MIB)—for a given α' can be computed as follows:

$$MDB = \arg \min_{\nabla_i} \mathcal{P}_{CD}(\nabla_i) > \tilde{\mathcal{P}}_{CD}, \forall i = 1, \dots, n \tag{43}$$

$$MIB = \arg \min_{\nabla_i} \mathcal{P}_{CI}(\nabla_i) > \tilde{\mathcal{P}}_{CI}, \forall i = 1, \dots, n \tag{44}$$

Equation (43) gives the smallest outlier ∇_i that leads to its detection for a given correct detection rate $\tilde{\mathcal{P}}_{CD}$, whereas (44) provides the smallest outlier ∇_i that leads to its identification for a given correct identification rate $\tilde{\mathcal{P}}_{CI}$.

As a consequence of the inequality in (36), the MIB will be larger than MDB, i.e., $MIB \geq MDB$. For the special case of having only one single alternative hypothesis, there is no difference between the MDB and MIB [46]. The computation of MDB_0 is easily performed by Equations (13) or (15), whereas the computation of the MDB in Equation (43) and the MIB in Equation (44) must be computed using Monte Carlo because the acceptance region (as well as the critical region) for the case of multiple alternative hypotheses is analytically intractable.

The non-centrality parameter for detection ($\lambda_{q=1}^{(MDB)}$) and identification ($\lambda_{q=1}^{(MIB)}$) for *IDS* can be computed similarly to Equation (12) as follows, respectively:

$$\lambda_{q=1}^{(MDB)} = \frac{MDB_{(i)}^2}{\sigma_{\nabla_i}^2} \tag{45}$$

$$\lambda_{q=1}^{(MIB)} = \frac{MIB_{(i)}^2}{\sigma_{\nabla_i}^2} \tag{46}$$

Thus,

$$\frac{MIB_{(i)}}{MDB_{(i)}} = \sqrt{\frac{\lambda_{q=1}^{(MIB)}}{\lambda_{q=1}^{(MDB)}}} \tag{47}$$

Note from Equation (47) that the relationship between the non-centrality parameters for detection ($\lambda_{q=1}^{(MDB)}$) and identification ($\lambda_{q=1}^{(MIB)}$) do not depend on the variance (or standard deviation) of estimated outlier $\sigma_{\nabla_i}^2$.

In the case of *IDS*, the power depends not only on the rate of Type II and Type III decision errors but also on the rate of over-identifications and the probability of statistical overlap. In the next section, we provide a procedure for computing the errors and success rates associated with *IDS*.

5. Computational Procedure for the Estimation of Success and Failure Rates of Iterative Outlier Elimination

After finding the critical value \hat{k} by the process described in Section 3.1, the procedure based on Monte Carlo is also applied to compute the probability levels of *IDS* when there is an outlier in the dataset as follows (summarised as a flowchart in Figure 1).

First, random error vectors are synthetically generated on the basis of a multivariate normal distribution because the assumed stochastic model for random errors is based on the matrix covariance of the observations. Here, we use the Mersenne Twister algorithm [72] to generate a sequence of random numbers and Box–Muller [73] to transform it into a normal distribution.

The magnitude intervals of simulated outliers are user-defined. The magnitude intervals are based on the standard deviation of the observation, e.g., $|3\sigma|$ to $|6\sigma|$, where σ is the standard deviation of the observations. Since the outlier can be positive or negative, the proposed algorithm randomly selects the signal of the outlier (for $q = 1$). Here, we use the discrete uniform distribution to select the signal of the outlier. Thus, the total error (ε) is a combination of random errors, and its corresponding outlier is as follows:

$$\varepsilon = e + c_i \nabla_i \quad (48)$$

In Equation (48), e is the random error generated from the normal distribution according to Equation (2), and the second part $c_i \nabla_i$ is the additional parameter that describes the alternative model according to Equation (8). Next, we compute the least-squares residuals vector according to Equation (3), but now we use the total error (ε) as follows:

$$\hat{e} = R\varepsilon \quad (49)$$

For *IDS*, the hypothesis of (8) for $q = 1$ (one outlier) is assumed, and the corresponding test statistic is computed according to (9). Then, the maximum test statistic value is computed according to Equation (16). Now, the decision rule is based on the critical value \hat{k} computed by Monte Carlo (see the steps (24)–(28) from Section 3.1). After identifying the measurement suspected to be the most likely outlier, it is excluded from the model, and least-squares estimation (LSE) and data snooping are applied iteratively until there are no further outliers identified in the dataset. Every time that a measurement suspected to be the most likely outlier is removed from the model, we check whether the normal matrix $A^T W A$ is invertible or not. If the determinant of $A^T W A$ is 0, $\det|A^T W A| = 0$, then there is a necessary and sufficient condition for a square matrix $A^T W A$ to be non-invertible. In other words, we check whether or not there is a solution available in the sense of ordinary LSE after removing a possible outlier.

The *IDS* procedure is performed for m experiments of random error vectors for each experiment contaminated by an outlier in the i th measurement. Therefore, for each measurement contaminated by an outlier, there are $v = 1, \dots, m$ experiments.

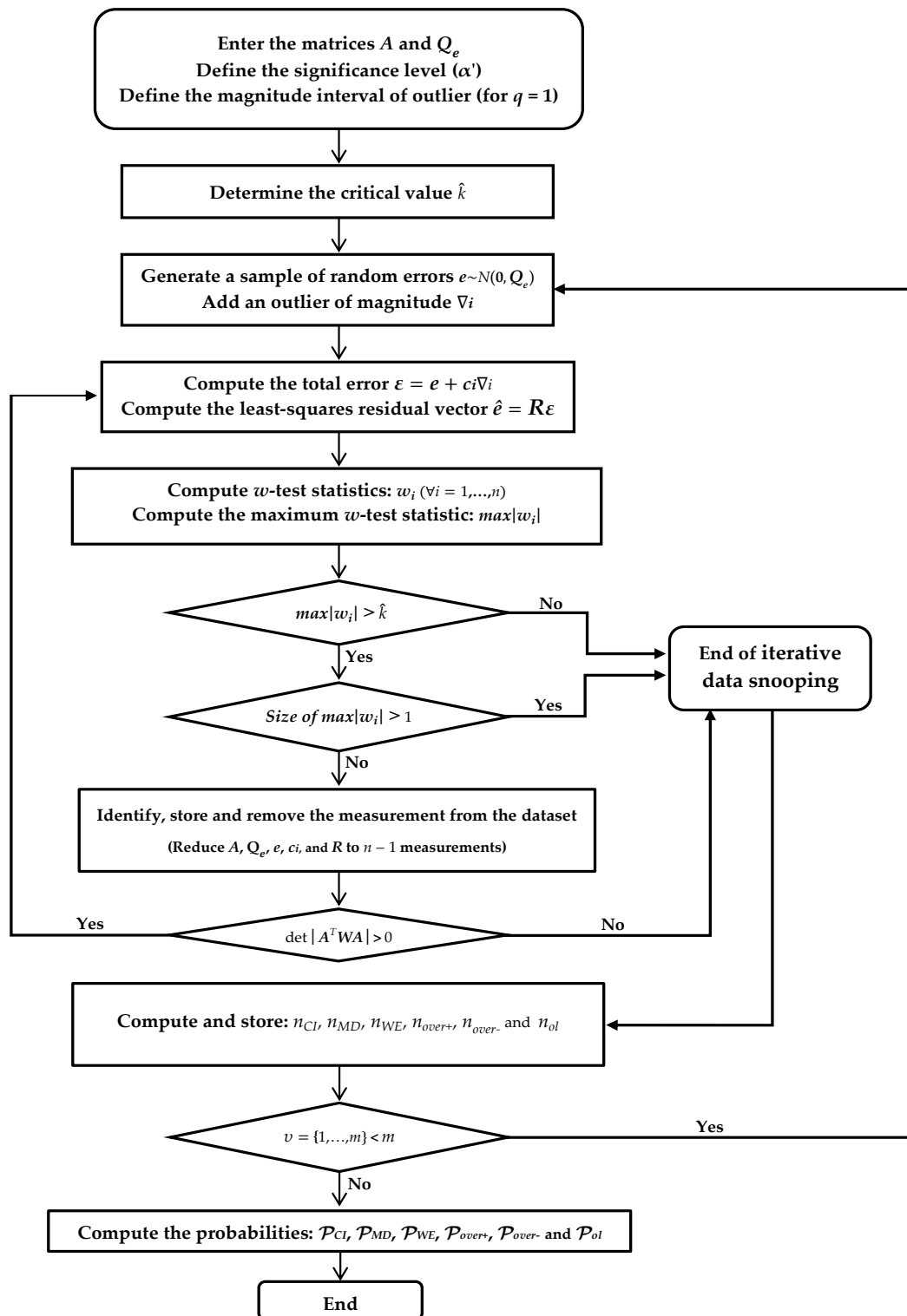


Figure 1. Flowchart of the algorithm to compute the probability levels of Iterative Data Snooping (IDS) for each measurement in the presence of an outlier.

The probability of correct identification \mathcal{P}_{CI} is obtained by the ratio between the correct identification cases and possible cases. Thus, if m is the total number of Monte Carlo experiments

(possible cases), then we count the number of times that the outlier is correctly identified (denoted as n_{CI}) and then approximate the probability of correct identification \mathcal{P}_{CI} as

$$\mathcal{P}_{CI} = \frac{n_{CI}}{m} \quad (50)$$

Similar to Equation (50), false decisions are computed as

$$\mathcal{P}_{MD} = \frac{n_{MD}}{m} \quad (51)$$

$$\mathcal{P}_{WE} = \frac{n_{WE}}{m} \quad (52)$$

$$\mathcal{P}_{over+} = \frac{n_{over+}}{m} \quad (53)$$

$$\mathcal{P}_{over-} = \frac{n_{over-}}{m} \quad (54)$$

$$\mathcal{P}_{ol} = \frac{n_{ol}}{m} \quad (55)$$

where:

- n_{MD} is the number of experiments in which *IDS* does not detect the outlier (\mathcal{P}_{MD} corresponds to the rate of missed detection);
- n_{WE} is the number of experiments in which the *IDS* procedure flags and removes only one single non-outlying measurement while the ‘true’ outlier remains in the dataset (\mathcal{P}_{WE} is the wrong exclusion rate);
- n_{over+} is the number of experiments in which *IDS* correctly identifies and removes the outlying measurement and others, and \mathcal{P}_{over+} corresponds to its probability;
- n_{over-} represents the number of experiments in which *IDS* identifies and removes more than one non-outlying measurement, whereas the ‘true outlier’ remains in the dataset (\mathcal{P}_{over-} is the probability corresponding to this error probability class); and
- n_{ol} is the number of experiments in which the detector in Equation (16) flags two (or more) measurements simultaneously during a given iteration of *IDS*. Here, this is referred to as the number of statistical overlap n_{ol} , and \mathcal{P}_{ol} corresponds to its probability.

In contrast to [12], in this paper, the probability levels associated with *IDS* are evaluated for each observation individually and for each outlier magnitude. Furthermore, we take care to control the *family-wise error rate*. In the next sections, we show the application of the algorithm described in Figure 1 to compute statistical quantities for *IDS*.

6. On the Probability Levels of Iterative Outlier Elimination

In this experiment, we considered two closed levelling networks: one with a low correlation between residuals and another one with a high correlation. For the low correlation case, we used the network given by Rofatto et al. [60], whereas for the high correlation, we chose the network from Knight et al. [43]. Figures 2a,b show the configuration of these networks, respectively.

Figure 3 shows an example of levelling for a single line. The equipment used to measure the level difference is an electronic digital level. In this case, the instrument is designed to operate by employing electronic digital image processing and is used in conjunction with a special barcoded staff (also called a barcode rod). After an operator accomplishes rough levelling with a bull’s eye bubble, an electronic digital level can be used to digitally obtain the barcoded staff reading. The result can be recorded manually in a levelling field-book or automatically stored in the data collector of the digital level. A simple height difference is obtained from the difference between the backsight staff reading and the foresight staff reading. An example of a “digital level – barcode staff” system is displayed in Figure 4. For more details about digital levels, see, for example [75–77].

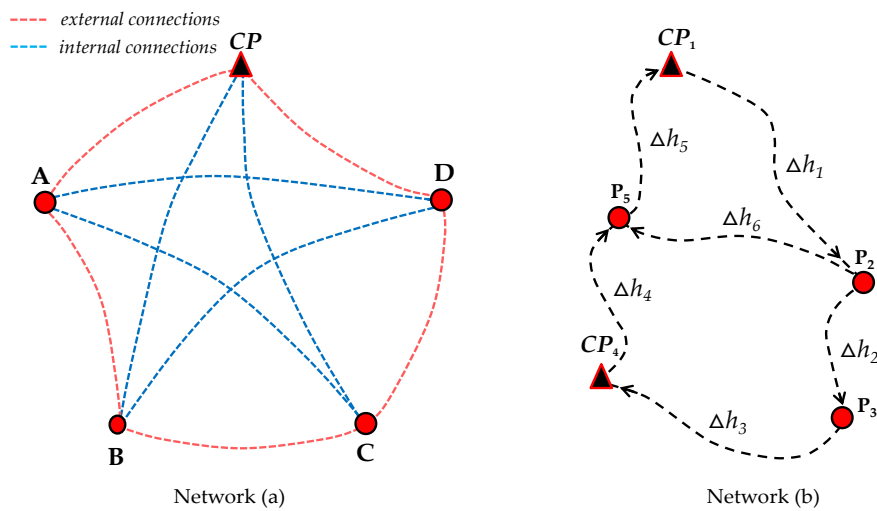


Figure 2. Levelling geodetic networks: (a) Levelling network adapted from [60] with a low correlation between w -test statistics; (b) Levelling network adapted from [43] with a high correlation between w -test statistics.

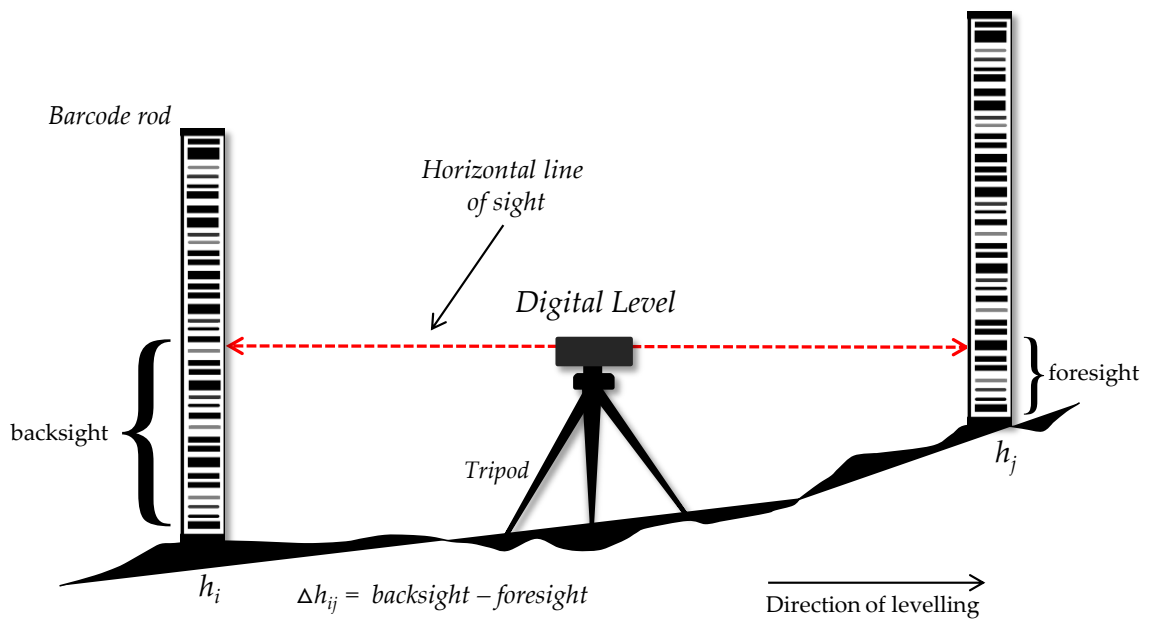


Figure 3. Example of levelling for a single line: the digital level instrument is placed between the points whose height difference is to be found. Special barcoded staffs are placed at these points and then sighted through the digital level instrument.



Figure 4. Example of a digital level – barcode staff system.

There are several levelling lines available for a levelling geodetic network. In the absence of outliers, i.e., under \mathcal{H}_0 , the model for levelling a geodetic network can be written in the sense of the standard Gauss–Markov model in Equation (1) as follows:

$$\Delta h_{i-j} + e_{\Delta h_{i-j}} = h_j - h_i, \quad (56)$$

where Δh_{i-j} is the height difference measured from point i to j , and $e_{\Delta h_{i-j}}$ is the random error associated with the levelling measurement. Generally, one of these points has a known height h , from which the height of another point is determined. The point with the known height is referred to here as the control point or benchmark (denoted by CP).

From network (a) in Figure 2, we have one control point and four points with unknown heights (A, B, C and D) for a total of four minimally constrained points. The control point is fixed (hard constraint or non-stochastic variable). In that case, matrix A is given by

$$A = \begin{pmatrix} -1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \quad (57)$$

In Figure 2, the red dashed lines are the external connections among the points of the network (a), whose measurements are Δh_{A-CP} , Δh_{A-B} , Δh_{B-C} , Δh_{C-D} and Δh_{D-CP} , whereas the blue dashed lines are the internal connections, whose measurements consist of Δh_{A-D} , Δh_{A-C} , Δh_{B-CP} , Δh_{B-D}

and Δh_{C-CP} . The distances of the external and internal connections are considered 240 m and 400 m, respectively. The equipment used is a spirit level with a nominal standard deviation for a single staff reading of 0.02 mm/m. Lines of sight distances are kept at 40 m. Each partial height difference, in turn, involves one instrument setup and two sightings: forward and back. Thus, the standard deviation for each total height difference equals

$$\begin{aligned}\sigma_{\Delta h_{i-j}} &= \sqrt{2p} \times 40 \text{ m} \times \frac{0.02 \text{ mm}}{m} \\ &= \sqrt{2p} \times 0.8 \text{ mm},\end{aligned}\quad (58)$$

where p is the number of partial height differences. In this case, each total height difference between external or internal connections is made of, respectively, three or five partial height differences. The readings are assumed to be uncorrelated and have equal uncertainty. In this case, the standard deviations of the measures for external and internal connections are 1.96 mm and 2.53 mm, respectively.

On the other hand, from network (b) in Figure 2, there are two control stations (fixed) and three user-stations with unknown heights. Matrix A is given by

$$A = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \\ -1 & 0 & 1 \end{pmatrix}\quad (59)$$

For network (b), we have the following measurements: $\Delta h_1 = \Delta h_{CP_1-P_2}$, $\Delta h_2 = \Delta h_{P_2-P_3}$, $\Delta h_3 = \Delta h_{P_3-CP_4}$, $\Delta h_4 = \Delta h_{CP_4-P_5}$, $\Delta h_5 = \Delta h_{P_5-CP_1}$ and $\Delta h_6 = \Delta h_{P_2-P_5}$. In this case, the covariance matrix of the measurements (metro units) is given by [43]

$$Q_e = \begin{pmatrix} 5.5 & 3.7 & 0.3 & -3.2 & -0.5 & 0.1 \\ 3.7 & 3.9 & 0.0 & -0.8 & -0.6 & -0.7 \\ 0.3 & 0.0 & 0.8 & -1.4 & 0.1 & 0.8 \\ -3.2 & -0.8 & -1.4 & 5.4 & -0.3 & -2.1 \\ -0.5 & -0.6 & 0.1 & -0.3 & 0.2 & 0.3 \\ 0.1 & -0.7 & 0.8 & -2.1 & 0.3 & 1.4 \end{pmatrix}\quad (60)$$

The correlation coefficients ρ_{w_i, w_j} between w -test statistics were computed for both network (a) and network (b) according to Equation (23). Table 2 provides the correlation ρ_{w_i, w_j} for network (a), and Table 3 provides it for network (b). In general, the correlation ρ_{w_i, w_j} for network (b) is much higher than that for network (a).

From Table 2, we observe that the maximum correlation is $\rho_{w_i, w_j} = \pm 0.4146$ for network (a) (i.e., $\rho_{w_i, w_j} = \pm 41.46\%$). In this case, as the correlation coefficient is less than 50%, the missed detection rate \mathcal{P}_{MD} is expected to be larger than the other decision errors of IDS.

From Table 3, it is expected that the wrong exclusion rate \mathcal{P}_{WE} will be significantly more pronounced than other wrong decisions of IDS. This is because of the high correlation between test statistics for network (b). Note also that the correlation coefficient between the second (Δh_2) and third Δh_3 measurement is exactly 1.00 (i.e., $\rho_{w_i, w_j} = 100\%$). This means that if one of these measurements is an outlier, then its corresponding w -test statistics will overlap. Therefore, an outlier can never be identified if it occurs in one of these measurements, but it can be detected.

Table 2. Correlation matrix of w -test statistics for levelling network (a).

Δh_{i-j}	A-CP	A-B	C-B	D-C	CP-D	D-A	C-A	CP-B	D-B	C-CP
A-CP	1.0000	-0.4146	-0.0488	-0.0488	-0.4146	-0.3464	-0.3134	-0.3464	-0.0660	-0.3134
A-B		1.0000	0.4146	0.0488	0.0488	-0.3134	-0.3464	0.3464	0.3134	0.0660
C-B			1.0000	0.4146	0.0488	-0.0660	-0.3464	-0.3134	-0.3464	0.3134
D-C				1.0000	0.4146	-0.3134	0.3134	-0.0660	-0.3464	-0.3464
CP-D					1.0000	0.3464	0.0660	-0.3134	0.3134	-0.3464
D-A						1.0000	-0.2565	-0.0223	-0.2565	0.0223
C-A							1.0000	0.0223	-0.0223	0.2565
CP-B								1.0000	-0.2565	-0.2565
D-B									1.0000	-0.0223
C-CP										1.0000

Table 3. Correlation matrix of w -test statistics for levelling network (b).

Δh_{i-j}	Δh_1	Δh_2	Δh_3	Δh_4	Δh_5	Δh_6
Δh_1	1.00	-0.41	-0.41	0.96	0.98	0.97
Δh_2		1.00	1.00	-0.36	-0.50	-0.61
Δh_3			1.00	-0.36	-0.50	-0.61
Δh_4				1.00	0.98	0.93
Δh_5					1.00	0.98
Δh_6						1.00

In the following subsections, we compute and analyse the probability levels associated with IDS for two cases. In the first part, the dataset is considered to be free of outliers, whereas, in the second one, there is an outlier in the dataset.

6.1. Scenario Coinciding with the Null Hypothesis: Absence of Outliers

In this context, we investigated the extent to which the Bonferroni correction deviates from the distribution of $\max-w$ on the basis of Monte Carlo. For this purpose, we considered the following Type I decision error rates: $\alpha' = 0.001$, $\alpha' = 0.0027$, $\alpha' = 0.01$, $\alpha' = 0.025$, $\alpha' = 0.05$ and $\alpha' = 0.1$. It is important to mention that the probability of a Type I Error of $\alpha' = 0.0027$ corresponds to a critical value of $k = 3$ in the case of a single test, which is known as the 3σ -rule [32]. We also set up $m = 200,000$ Monte Carlo experiments, as proposed by [12]. For each α' , we computed the corresponding critical value according to Bonferroni from Equation (22) and on the basis of Monte Carlo from the procedure described in Section 3, specifically from step (24) to (28). Both methods were applied for networks (a) and (b) in Figure 2. The result is displayed in Figure 5.

From Figure 5, we observe that the critical values computed from Monte Carlo are always smaller than those values computed by Equation (22) for both networks. This is because matrix R in Equation (3) promotes the correlation between residuals. Note that matrix R depends on the network geometry given by matrix A . This means that we will always have some degree of correlation between residuals. If we neglect the correlation between residuals, as in (22), then we are assuming that there is no association between residuals. Thus, we overestimate the probability of $\max-w$ by using (22), and the dashed curve in Figure 5 is always above the solid curve. Therefore, with Bonferroni, the user has no full control over Type I Errors. We point out some particularities as follows.

The Bonferroni method can only be used with a good approximation to control the Type I Error of IDS for the case in which we have a measurement system with high redundancy and low residual correlation (i.e., low correlation between w -test statistics ρ_{w_i, w_j}) and for small α' . This is the case for network (a). On the other hand, Bonferroni does not work well for network (b). In the latter case, the effect is more pronounced because of the low redundancy and high residual correlation. In general, under the condition that the correct decision is made when the null hypothesis \mathcal{H}_0 is accepted, it is less likely to get a large $\max-w$ than the prediction by (22).

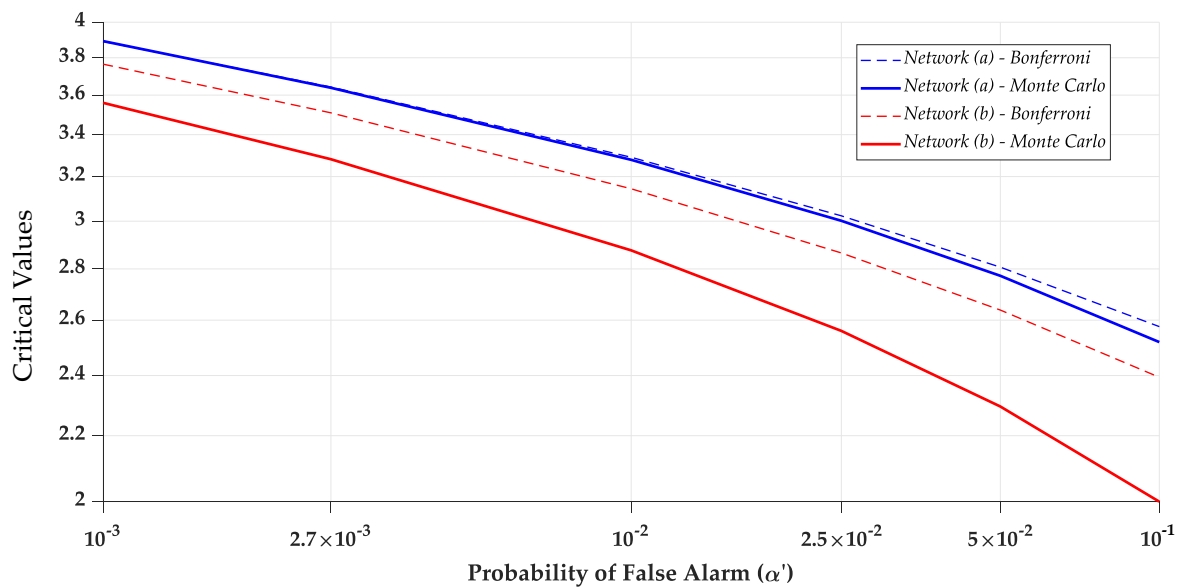


Figure 5. Critical values of $\max-w$ for both levelling networks (a) in blue and (b) in red. Solid curves were obtained from the Monte Carlo procedure in Section 4 with $m = 200,000$ Monte Carlo experiments. Dashed curves were obtained from Bonferroni in (22).

Here, we treated the extreme (i.e., maximum absolute) normalised residuals $\max-w$ in (16) directly as a test statistic. Figure 6 shows the distribution of $\max-w$ for both networks (a) and (b). We observe that the distribution of $\max-w$ for network (a) gets closer to a normal distribution than network (b). This means that the smaller the correlation between residuals, the smaller the Type I decision error rate α' and, therefore, the larger the critical values. Figure 7 shows the cumulative distribution of $\max-w$ for both networks (a) and (b). Figure 7 reveals that, for the same level of error probability, the critical value for network (a) is always smaller than the one computed for network (b).

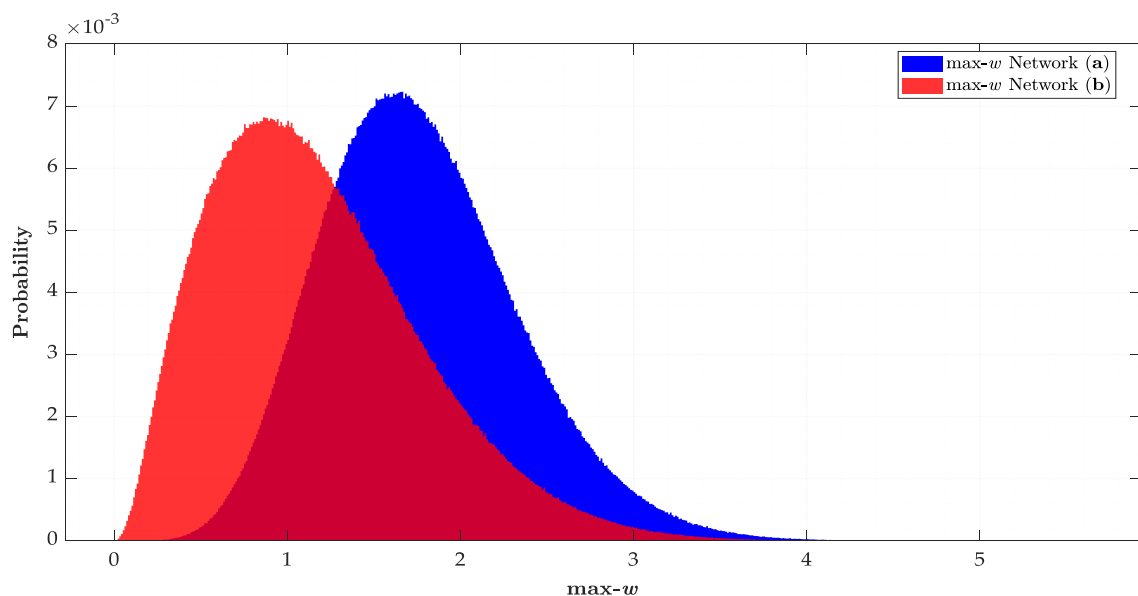


Figure 6. Probability histograms of $\max-w$ for networks (a) in blue and (b) in red.

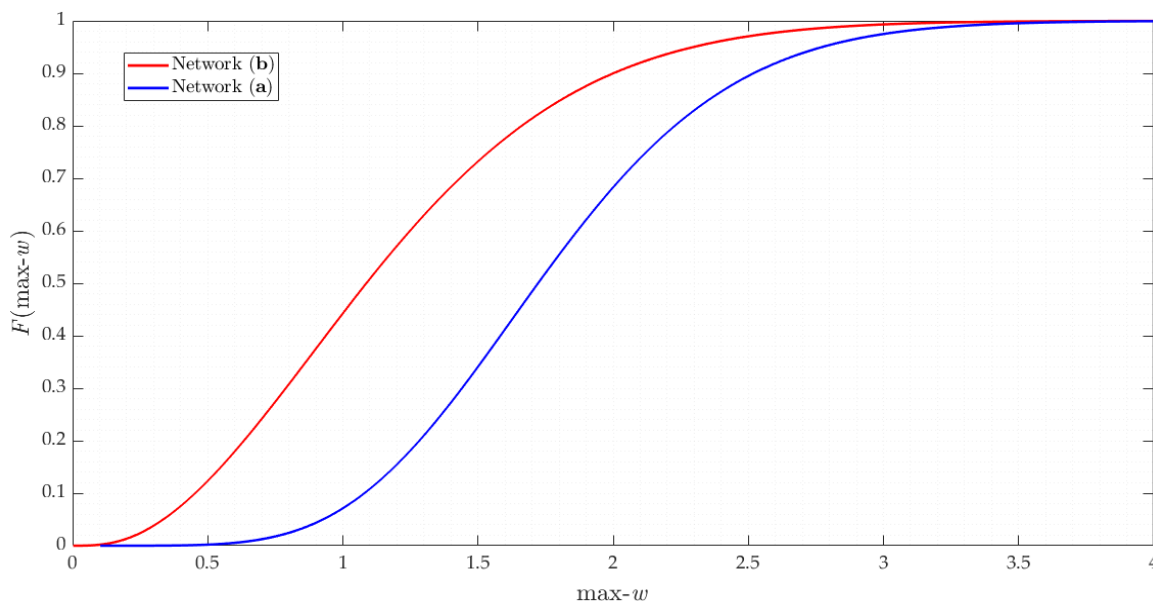


Figure 7. Cumulative frequency of max-w for networks (a) in blue and (b) in red.

Until now, our outcomes have been investigated under the condition that \mathcal{H}_0 is true. In the next subsection, we analyse the probability levels of *IDS* in the presence of an outlier. In that case, the decision rule is based on critical values from the max-w distribution, as detailed in Table 4. It is important to note that the critical value 3 of the 3σ -rule is not valid for a multiple test case. In fact, the critical values for outlier identification (i.e., a multiple test) depend on the geometry of the network and the sensor uncertainty. Therefore, the probability associated with the 3σ -rule for network (a) will be close to $\alpha' = 0.025$, and that for network (b) will be close to $\alpha' = 0.0067$.

Table 4. Critical values from the Bonferroni and Monte Carlo procedures for networks (a) and (b).

α'	k_{bonf} from (22) for Net (a)	k_{bonf} from (22) for Net (b)	\hat{k} from (28) for Net (a)	\hat{k} from (28) for Net (b)
0.001	3.89	3.76	3.89	3.56
0.0027	3.64	3.51	3.64	3.28
0.01	3.29	3.14	3.28	2.88
0.025	3.02	2.87	3	2.56
0.05	2.81	2.64	2.77	2.29
0.1	2.58	2.39	2.52	2

6.2. Scenario Coinciding with an Alternative Hypothesis: Presence of an Outlier

In this scenario, there is an outlier in the dataset. Thus, the correct decision is made when the alternative hypothesis $\mathcal{H}_A^{(i)}$ from Equation (8) is accepted. In this step, we computed the probability levels of *IDS* using the procedure in Section 5. The decision rule is based on critical values computed by Monte Carlo. These values are presented in Table 4. We arbitrarily defined the outlier magnitude from $|3\sigma|$ to $|8\sigma|$ for network (a) and $|1\sigma|$ to $|12\sigma|$ for network (b).

The sensitivity indicators MDB and MIB were also computed according to Equations (43) and (44), respectively. The success rates for outlier detection and outlier identification were taken as equal to 0.8, i.e., $\tilde{\mathcal{P}}_{CD} = \tilde{\mathcal{P}}_{CI} = 0.8$, respectively.

6.2.1. Geodetic Network with Low Correlation between Residuals

We start from network (a) with a low correlation between residuals. We observe that there is a high degree of homogeneity for network (a). This can be explained by the redundancy numbers, denoted by (r_i) . The redundancy numbers are the elements of the main diagonal of matrix \mathbf{R} in Equation (3). The redundancy number (r_i) is an internal reliability measure that represents the ability

of a measurement to project the measurement error in the least-squares residuals. Then, the higher the number of redundancy, the higher the resistance of the measurement to outliers.

The redundancy numbers for the measurements constituting external connections are identical and equal to $r_i = 0.519$, whereas the measurements constituting external connections are also identical but equal to $r_i = 0.681$. Consequently, the probability levels associated with *IDS* are practically identical for both external and internal connections. Thus, we subdivided our result into two parts: mean values of the probability levels for external connections and mean values of the probability levels for internal connections.

Figure 8 shows the probability of correct identification (\mathcal{P}_{CI}) and correct detection (\mathcal{P}_{CD}) in the presence of an outlier for both external and internal connections. In general, we observe that the larger the Type I Error α' (or the lower the critical value \hat{k}), the higher the rate of correct detection \mathcal{P}_{CD} . This is not fully true for outlier identification \mathcal{P}_{CI} .

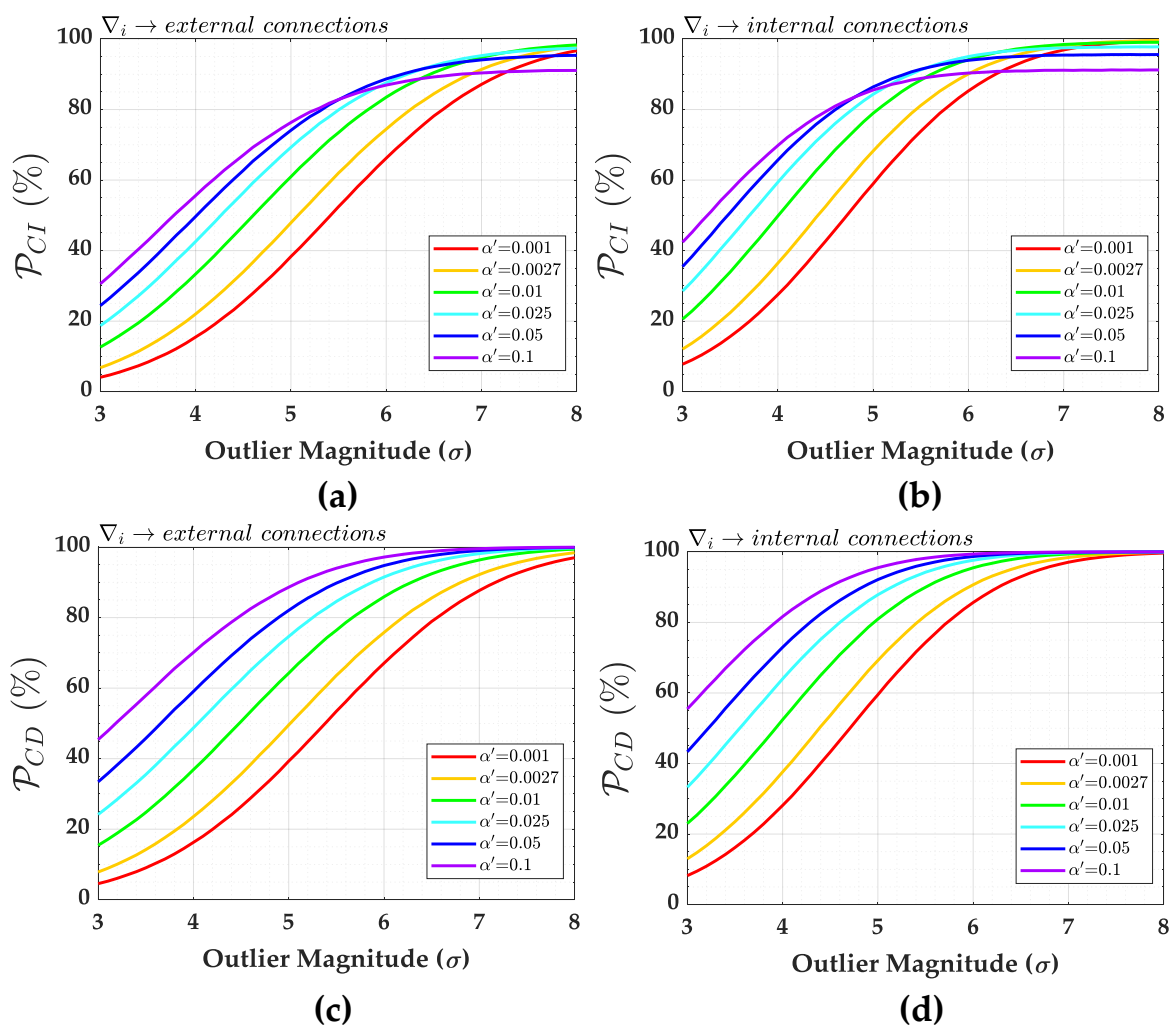


Figure 8. Probability of correct identification (\mathcal{P}_{CI}) and correct detection (\mathcal{P}_{CD}) for network (a): (a) \mathcal{P}_{CI} for the measurements constituting external connections. (b) \mathcal{P}_{CI} for the measurements constituting internal connections. (c) \mathcal{P}_{CD} for the measurements constituting external connections. (d) \mathcal{P}_{CD} for the measurements constituting internal connections.

We observe that the probability of correct identification \mathcal{P}_{CI} becomes constant from a certain outlier magnitude. Moreover, the larger the α' , the faster the success rate at which outlier identification \mathcal{P}_{CI} stabilizes. In other words, the larger the α' , the higher the \mathcal{P}_{CI} , but only up to a certain limit of outlier magnitude. After this bound, there is an inversion: the larger the α' , the lower the probability

of correct identification \mathcal{P}_{CI} . This can be explained by the following: (i) the larger the α' , the larger the critical region (or the smaller the acceptance region) of the working hypothesis \mathcal{H}_0 ; (ii) the larger the critical region, the smaller the size of the test; (iii) the smaller the size of the test, the less likely the hypothesis test will identify a small difference. In other words, there is no significant difference among the probabilities of correct identification \mathcal{P}_{CI} for outliers lying within a certain location of the critical region. Therefore, the probabilities of correct identification \mathcal{P}_{CI} for those outliers are practically identical.

Let us take the probability of correct identification \mathcal{P}_{CI} for external connections in Figure 8 as an example. If the user chooses $\alpha' = 0.1$, then the test will be limited to 90% of the acceptance region. In this case, an outlier of 6.6σ will have a practically identical probability of correct identification of $\mathcal{P}_{CI} = 90\%$ of an outlier of 8σ (or greater than 8σ). However, if one chooses an $\alpha' = 0.001$ (99.9% of acceptance region), then a 6.6σ outlier would not be identified at the same rate as an 8σ outlier. Therefore, in that case, the Type I decision error α' (or the critical value \hat{k}) restricts the maximum rate of correct outlier identification \mathcal{P}_{CI} .

Note also that there are no significant differences between detection \mathcal{P}_{CD} and identification \mathcal{P}_{CI} rates for small Type I decision errors (see, e.g., $\alpha' = 0.001$ and $\alpha' = 0.0027$).

Furthermore, the probabilities of correct detection \mathcal{P}_{CD} and identification \mathcal{P}_{CI} are greater for internal than external connections. For an outlier of 4.5σ and $\alpha' = 0.1$, for instance, the probability of correct identification is $\mathcal{P}_{CI} = 67\%$ for external connections, whereas, for internal connections, it is $\mathcal{P}_{CI} = 80\%$.

Next, we compared the sensitivity indicators MDB and MIB by considering a success rate of 0.8 (80%) for both outlier identification and outlier detection, i.e., $\tilde{\mathcal{P}}_{CI} = \tilde{\mathcal{P}}_{CD} = 80\%$ (see Equations (43) and (44)). The user can also find the MDB and MIB for other success rates. The result is displayed in Figure 9. We observe that the larger the Type I decision error α' , the more that the MDB deviates from the MIB. In other words, the MIB stabilizes for a certain α' , whereas the MDB continues to decrease. It is harder to identify than it is to detect an outlier. Therefore, the MIB will always be greater than or equal to the MDB.

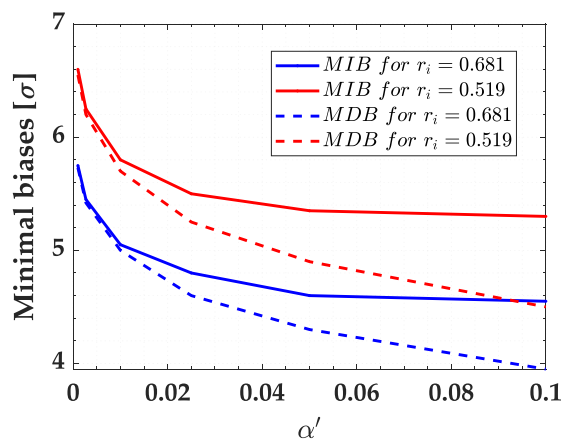


Figure 9. Minimal Detectable Bias (MDB) and Minimal Identifiable Bias (MIB) for both external and internal connections and for each α' .

The standard deviations of estimated outlier σ_{∇_i} for external and internal connection measurements are 2.7 mm and 3 mm, respectively. These σ_{∇_i} values were obtained by means of the square-root in Equation (14). Note from Equations (45) and (46) that the higher the accuracy of the outlier estimate, the lower the MDB and MIB, respectively. However, note from Equation (47) that the relationship between the MIB and MDB does not depend on $\sigma_{\nabla_i}^2$. This is true when the outlier is treated as bias. In other words, if outliers are treated as bias, then they act like systematic errors

by shifting the random error distribution by their own value [13]. The result for $\tilde{\mathcal{P}}_{CI} = \tilde{\mathcal{P}}_{CD} = 0.8$ is summarised in Table 5.

As can be seen from Table 5, in general, the MIB does not deviate too much from the MDB. This is because of a low correlation between residuals. The difference becomes larger when the Type I decision error α' is increased. Note, for instance, that the MDB and MIB are practically identical for Type I decision errors of $\alpha' = 0.001$ and $\alpha' = 0.0027$. In other words, an outlier is detected and identified with the same probability level when there is a low correlation between residuals and for small α' . Therefore, we observe that the larger the α' , the greater the difference between the MIB and MDB. In this case, the difference between the MIB and MDB is governed by the user-defined α' .

Table 5. Relationship between the MDB and MIB for network (a) by considering $\tilde{\mathcal{P}}_{CI} = \tilde{\mathcal{P}}_{CD} = 0.8$.

α'	(External Connections)			(Internal Connections)		
	$\lambda_{q=1}^{(MDB)}$	$\lambda_{q=1}^{(MIB)}$	MIB/MDB	$\lambda_{q=1}^{(MDB)}$	$\lambda_{q=1}^{(MIB)}$	MIB/MDB
0.001	22.27	22.61	1.01	22.36	22.52	1.00
0.0027	19.95	20.27	1.01	20.01	20.23	1.01
0.01	16.86	17.46	1.02	17.03	17.37	1.01
0.025	14.3	15.7	1.05	14.41	15.69	1.04
0.05	12.46	14.85	1.09	12.59	14.41	1.07
0.1	10.51	14.58	1.18	10.63	14.10	1.15

From Table 6, it can also be noted that the MIB is higher for internal than external connections. This is because internal connections are less precise than external connections. Therefore, the effect on the heights (model parameters) of an unidentified outlier is greater if the outlier magnitude is equal to the MIB of the internal connections. However, from Figure 8, we observe that it would be easier to identify an outlier if it occurred in the measurements that constitute internal connections than if it occurred in external connections.

Table 6. MIBs for each α' and for $\tilde{\mathcal{P}}_{CI} = 80\%$.

α'	MIB (m)	MIB (m)
	(External Connections)	(Internal Connections)
0.001	0.0129	0.0145
0.0027	0.0122	0.0138
0.01	0.0114	0.0128
0.025	0.0108	0.0121
0.05	0.0105	0.0116
0.1	0.0104	0.0115

It is important to mention that both the MDB and MIB are ‘invariant’ with respect to the control point position CP . This is a well-known fact and can already follow from the MDB and MIB definitions in Equations (45) and (46), respectively, which show that both the MDB and MIB are driven by the variance matrices of the measurements and adjusted residuals.

Figure 10 provides the result for the Type III decision error (\mathcal{P}_{WE}). In the worst case, we have $\mathcal{P}_{WE} = 0.12$ (12%) for $|3\sigma|$. In general, \mathcal{P}_{WE} is larger for external than internal connections. This is linked to the fact that the residual correlation ρ_{w_i, w_j} in Table 2 is higher for external than internal connections. Furthermore, the larger the Type I error rate α' , the larger the \mathcal{P}_{WE} for both internal and external connections. Because of the low probability of \mathcal{P}_{WE} decision errors for network (a), the user may opt for a larger α' so that the Type II decision error \mathcal{P}_{MD} is as small as possible. Thus, it is possible to guarantee a high outlier identification rate. This kind of analysis can be performed, for instance, during the design stage of a geodetic network (see, e.g., [60]).

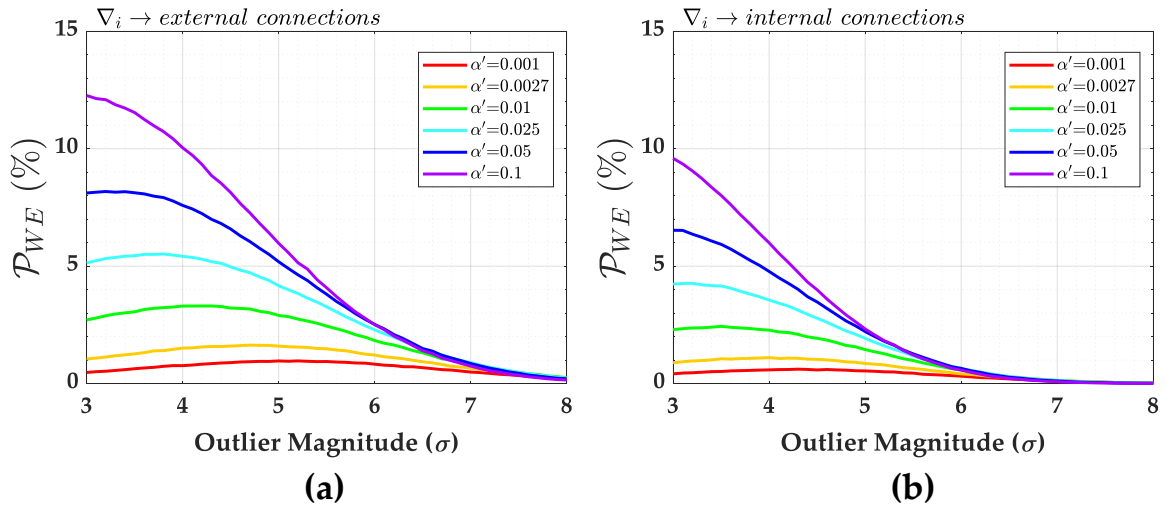


Figure 10. Probability of committing a Type III decision error (\mathcal{P}_{WE}) for network (a): (a) \mathcal{P}_{WE} for external connections. (b) \mathcal{P}_{WE} for internal connections.

Figure 10 gives only the overall rate of \mathcal{P}_{WE} . Figure 11, on the other hand, displays the individual contributions to \mathcal{P}_{WE} according to Equation (40) for $\alpha' = 0.1$. As expected, the higher the correlation coefficient between w -test statistics ρ_{w_i,w_j} , the greater the contribution of the measurement to \mathcal{P}_{WE} (see, e.g., [2]). In that case, we can also verify from Figure 12 that the larger the redundancy number r_i , the smaller the \mathcal{P}_{WE} . Moreover, the larger the outlier magnitude, the smaller the \mathcal{P}_{WE} . We also observe from Figure 13 that the larger the ρ_{w_i,w_j} , the larger the weighting factor $p_{i(\mathcal{P}_{WE})}$. The weighting factors $p_{i(\mathcal{P}_{WE})}$ for the highest correlations (i.e., $\rho_{w_i,w_j} = 0.415$ for $r_i = 0.519$ and $\rho_{w_i,w_j} = 0.346$ for $r_i = 0.681$) increase as the outlier magnitude increases. However, this is not significant. While the weighting factor $p_{i(\mathcal{P}_{WE})}$ for the highest correlation coefficient increases by around 1%, the overall \mathcal{P}_{WE} decreases by around 20%. In general, the weighting factor $p_{i(\mathcal{P}_{WE})}$ is relatively constant. The weighting factor $p_{i(\mathcal{P}_{WE})}$ was obtained by Equation (41).

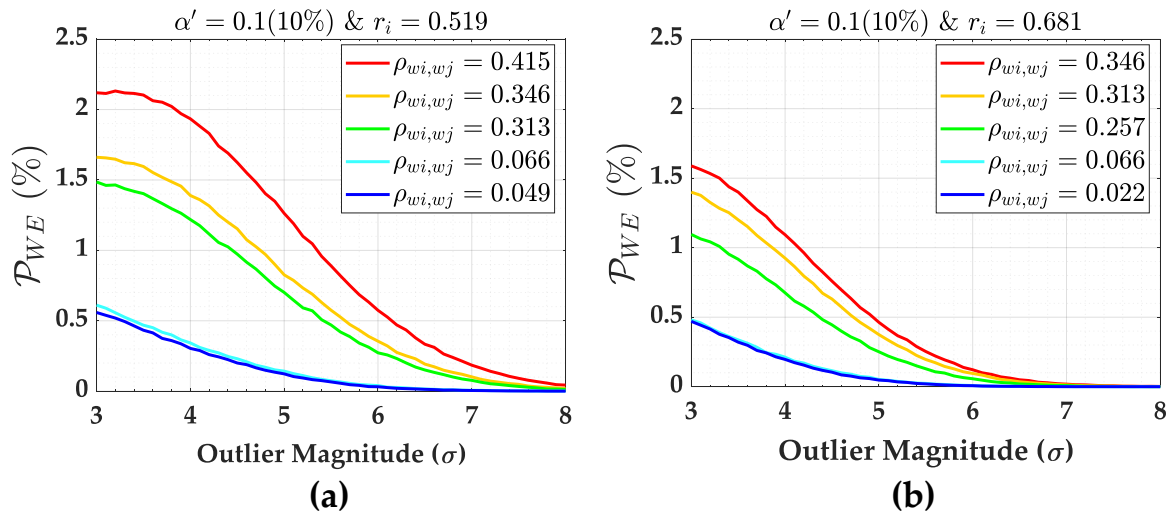


Figure 11. Individual contribution of external and internal connections to the overall \mathcal{P}_{WE} and their relationship with the correlation coefficient ρ_{w_i,w_j} for network (a). (a) Individual contribution to \mathcal{P}_{WE} by external connections. (b) Individual contribution to \mathcal{P}_{WE} by internal connections.

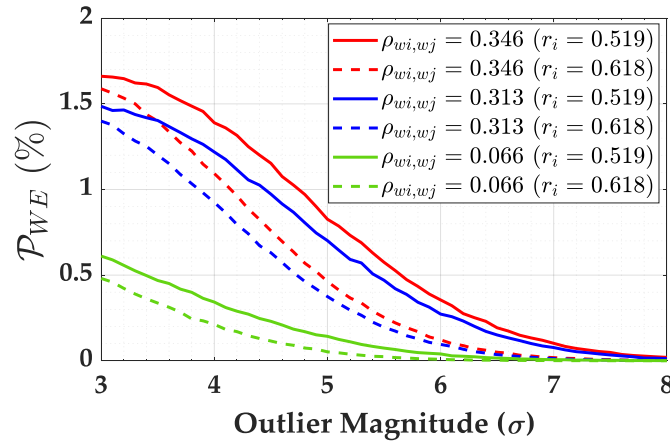


Figure 12. Individual contributions to the overall \mathcal{P}_{WE} for network (a) and their relationship with the redundancy number r_i for residuals with the same correlation coefficient ρ_{w_i,w_j} and $\alpha' = 0.1$.

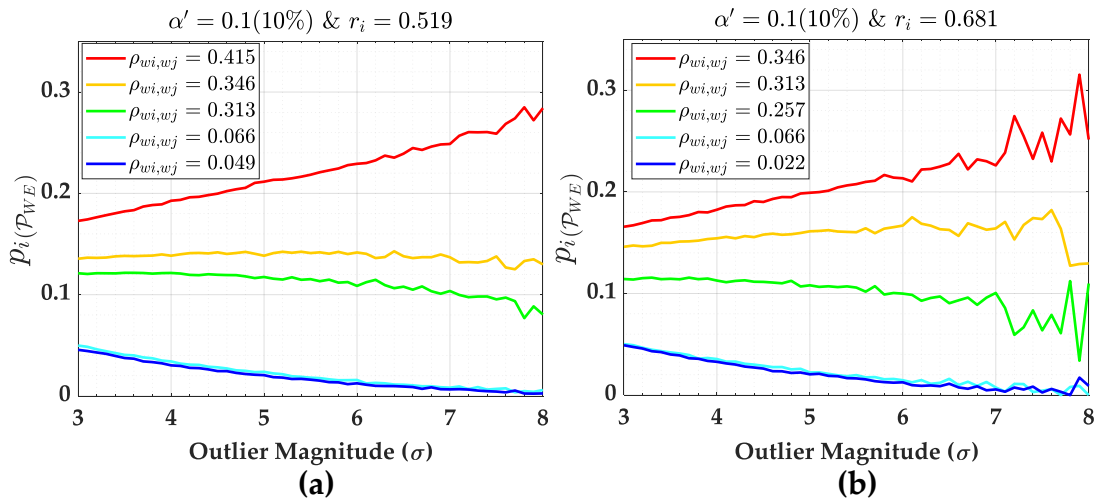


Figure 13. Weighting factors $p_{i(\mathcal{P}_{WE})}$ for network (a) and for $\alpha' = 0.1$. (a) Weighting factors $p_{i(\mathcal{P}_{WE})}$ for external connections. (b) Weighting factors $p_{i(\mathcal{P}_{WE})}$ for internal connections.

The over-identification cases \mathcal{P}_{over+} and \mathcal{P}_{over-} are presented in Figure 14. In general, the larger the Type I decision error α' , the larger the over-identification cases for that network. The larger the magnitude of the outlier, the larger the \mathcal{P}_{over+} and smaller the \mathcal{P}_{over-} . For small α' , we observe that \mathcal{P}_{over-} and \mathcal{P}_{over+} are practically null (see, e.g., for $\alpha' = 0.001$ and $\alpha' = 0.0027$). In general, the larger the correlation coefficient ρ_{w_i,w_j} , the smaller the \mathcal{P}_{over+} and the larger the \mathcal{P}_{over-} . Moreover, we also observe that the larger the redundancy number r_i , the larger the \mathcal{P}_{over+} and the smaller the \mathcal{P}_{over-} .

The probability of statistical overlap \mathcal{P}_{ol} is practically null for this network. This is because each point of network (a) has at least four connections. This means that even with an exclusion, there are still three measurement levels per point (i.e., three connections per point), which guarantees the minimum redundancy necessary for the second round of IDS. The very low residual correlation of this network also contributes to the non-occurrence of statistical overlap.

The results presented so far are valid for the case of a system with high redundancy and low residual correlation. In the next section, we present the results for a system with low redundancy and high residual correlation.

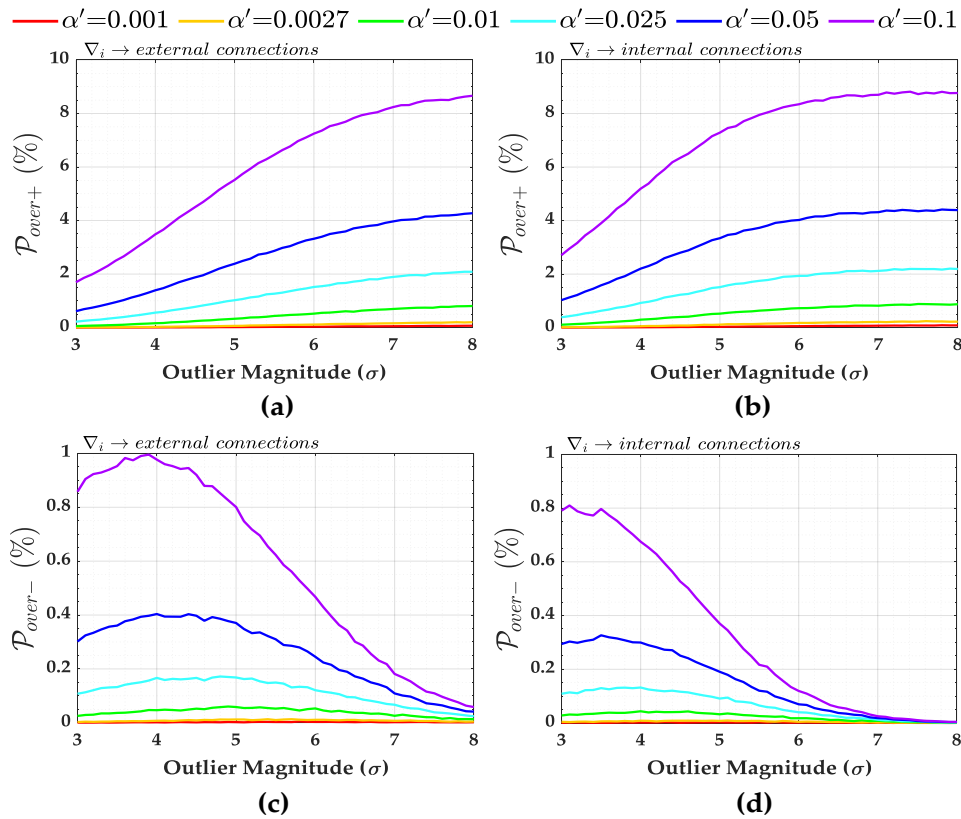


Figure 14. Over-identification cases for network (a). (a) \mathcal{P}_{over+} for external connections. (b) \mathcal{P}_{over+} for internal connections. (c) \mathcal{P}_{over-} for external connections. (d) \mathcal{P}_{over-} for internal connections.

6.2.2. Geodetic Network with High Correlation Between Residuals

Now, the correlation between residuals is very high. This is the case for network (b) detailed in Figure 2. Since the measurements are correlated for network (b), instead of redundancy numbers, reliability numbers (\bar{r}_i) should be given as an internal reliability measure, as follows [43]:

$$\bar{r}_i = c_i^T Q_e c_i c_i^T W Q_e W c_i, \forall i = 1, \dots, n \tag{61}$$

The reliability numbers (\bar{r}_i) in Equation (61) are equivalent to redundancy numbers when it is assumed that the measurements are uncorrelated. Table 7 gives the reliability numbers (\bar{r}_i), the standard deviation of each measurement $\sigma_{\Delta h_{i-j}}$ and the standard deviation of each estimated outlier σ_{∇_i} for network (b).

Table 7. Reliability numbers (\bar{r}_i), standard deviation $\sigma_{\Delta h_{i-j}}$ and standard deviation of estimated outlier σ_{∇_i} for network (b).

Δh_{i-j}	\bar{r}_i	$\sigma_{\Delta h_{i-j}}$ (m)	σ_{∇_i} (m)
Δh_1	10.58	2.35	0.72
Δh_2	0.62	1.97	2.50
Δh_3	0.13	0.89	2.50
Δh_4	13.68	2.32	0.63
Δh_5	1.95	0.45	0.32
Δh_6	3.56	1.18	0.63

The probabilities of correct identification (\mathcal{P}_{CI}) for this network are displayed in Figure 15. The critical values (\hat{k}) for network (b) are those given in Table 4. The probability levels of correct detection (\mathcal{P}_{CD}) are provided in Figure 16.

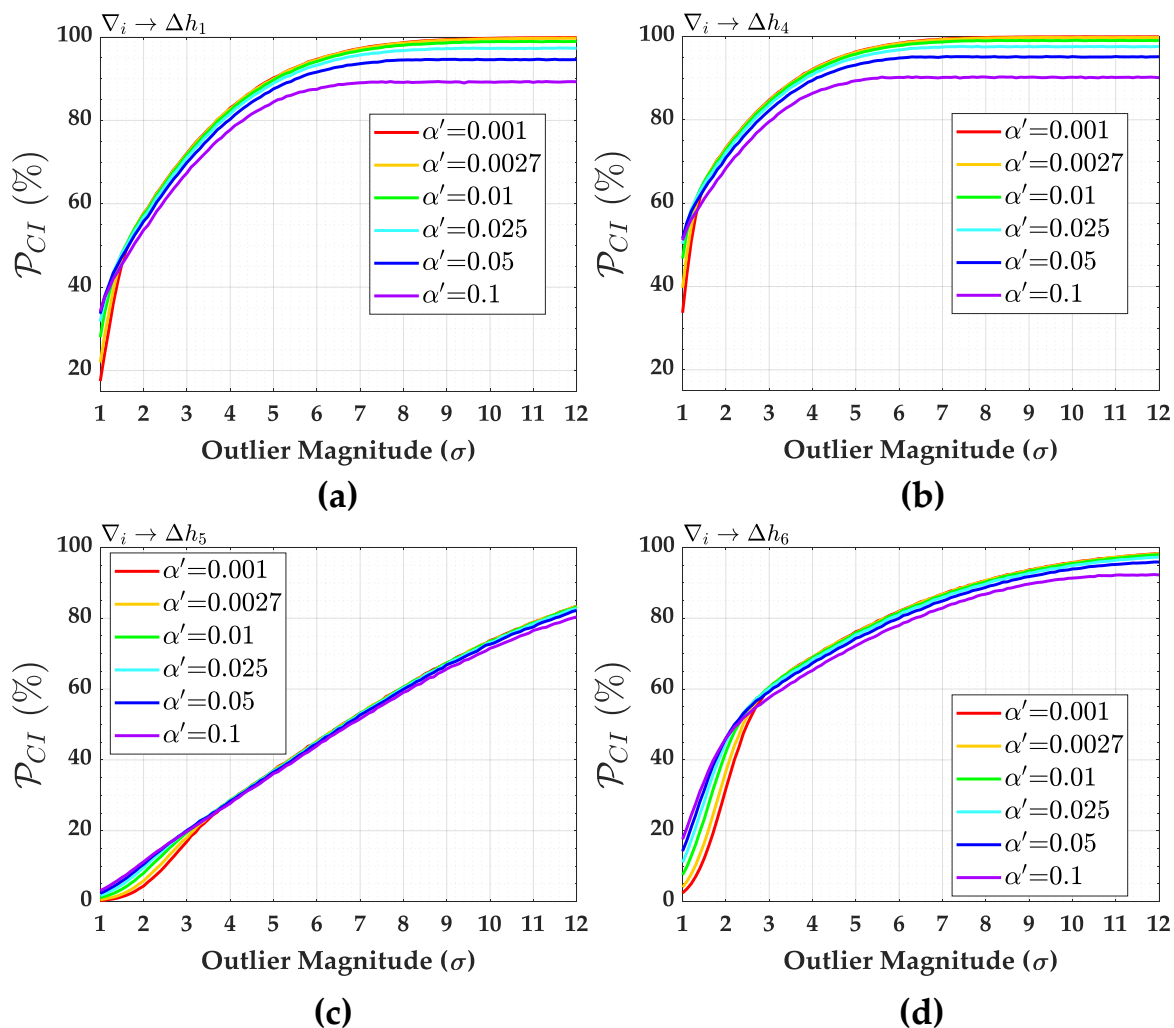


Figure 15. Probability of correct identification (\mathcal{P}_{CI}) for network (b). (a) \mathcal{P}_{CI} for Δh_1 . (b) \mathcal{P}_{CI} for Δh_4 . (c) \mathcal{P}_{CI} for Δh_5 . (d) \mathcal{P}_{CI} for Δh_6 .

In contrast to network (a), the probability of correct identification (\mathcal{P}_{CI}) for network (b) is different for each measurement. It is also found that the larger the Type I decision error α' , the higher the probability of correct identification (\mathcal{P}_{CI}). However, it is only true up to a certain level of outlier magnitude. After this magnitude level, the larger the Type I decision error α' , the lower the probability of correct identification (\mathcal{P}_{CI}).

The user-defined Type I error α' has indeed become less significant at a certain outlier magnitude. Note, for example, that the probability of correct identification for measurement Δh_1 for $\alpha' = 0.1$ is higher than that for $\alpha' = 0.001$ when the outlier magnitude is between 1σ and 1.5σ . For a magnitude greater than 1.5σ , we note that the larger the Type I decision error α' , the lower the probability of correct identification \mathcal{P}_{CI} . The choice of Type I error α' , however, has no significant effect on the probability of correct identification \mathcal{P}_{CI} for an outlier magnitude greater than 1.5σ . This analysis can also be done with Δh_4 , Δh_5 and Δh_6 .

There is no probability of identification for both measurements Δh_2 and Δh_3 . This is because the residual correlation of these measurements is equal to exactly one (i.e., $\rho_{w_i, w_j} = 1.00$). Furthermore, the reliability numbers (\bar{r}_i) in Table 7 for those measurements are close to zero. However, if one of those measurements were affected by a single outlier, then IDS would have the ability to detect it. In other words, there is reliability in terms of outlier detection for Δh_2 and Δh_3 .

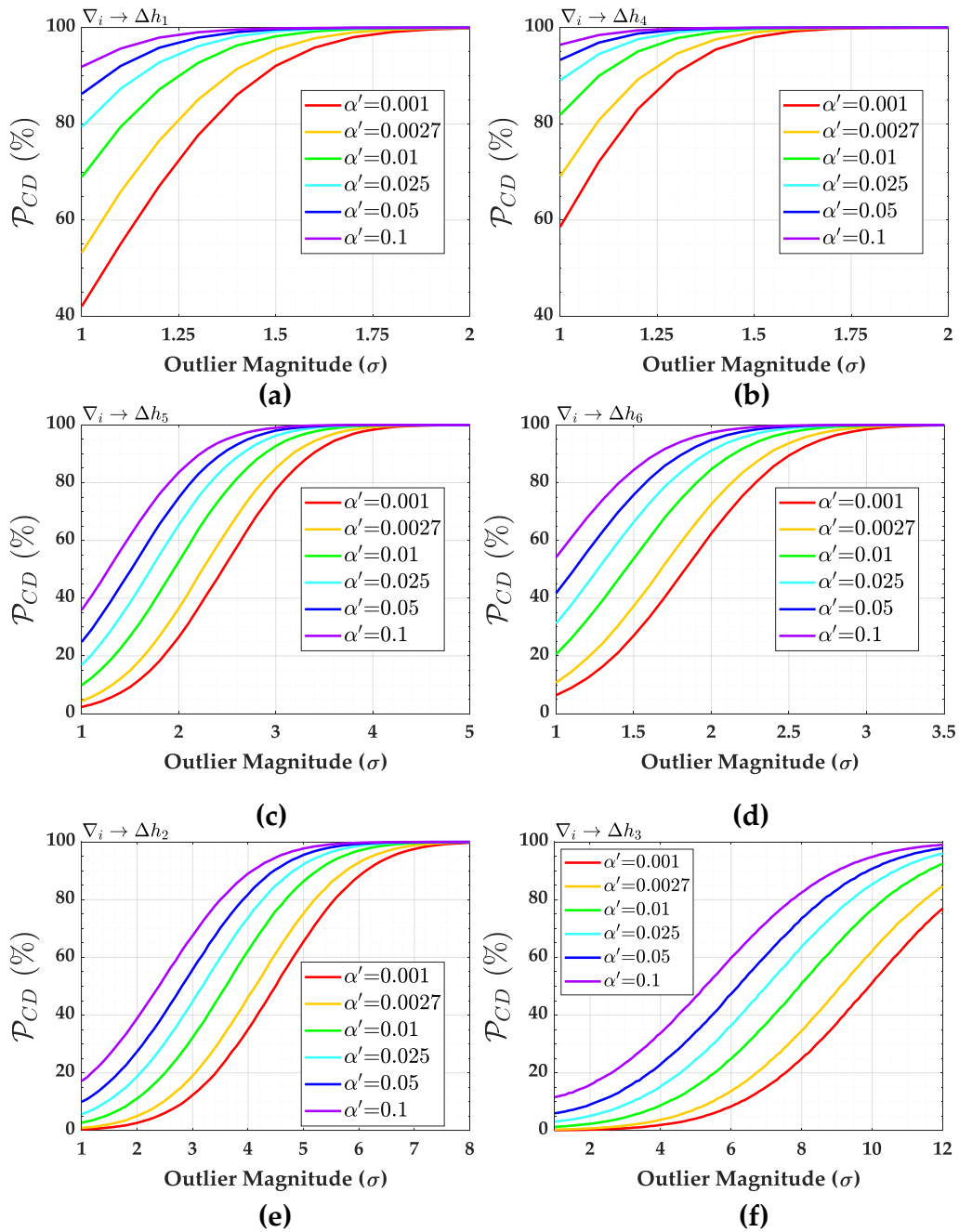


Figure 16. Probability of correct detection (\mathcal{P}_{CD}) for network (b). (a) \mathcal{P}_{CD} for Δh_1 . (b) \mathcal{P}_{CD} for Δh_4 . (c) \mathcal{P}_{CD} for Δh_5 . (d) \mathcal{P}_{CD} for Δh_6 . (e) \mathcal{P}_{CD} for Δh_2 . (f) \mathcal{P}_{CD} for Δh_3 .

We observe that the higher the reliability numbers in Table 3, the higher the power of detection \mathcal{P}_{CD} and identification \mathcal{P}_{CI} . In general, the larger the Type I decision error α' , the lower the probability of missed detection \mathcal{P}_{MD} and, therefore, the higher the probability of correct detection \mathcal{P}_{CD} .

The sensitivity indicators MDB and MIB for Δh_1 , Δh_4 , Δh_5 and Δh_6 are shown in Tables 8–11, respectively. Both MIBs and MDBs were computed for each α' and for a success rate of 0.8 (80%) for both outlier detection and identification, i.e., $\tilde{\mathcal{P}}_{CD} = \tilde{\mathcal{P}}_{CI} = 80\%$. The non-centrality parameters for outlier detection and identification were computed according to Equations (45) and (46), respectively. In general, the larger the Type I decision error α' , the larger the MIB and the smaller the MDB. In other words, the larger the Type I decision error α' , the greater the chances of outlier detection but the lower

the chances of outlier identification. In that case, the larger the Type I Error α' , the larger the MIB/MDB ratio. Therefore, an outlier with a size of the MDB should be enlarged in order to identify it [1,2,12,46].

Table 8. Relationship between the MDB and MIB for Δh_1 and for $\tilde{\mathcal{P}}_{CD} = \tilde{\mathcal{P}}_{CI} = 80\%$.

α'	MIB	MDB	$\lambda_{q=1}^{(MIB)}$	$\lambda_{q=1}^{(MDB)}$	MIB/MDB
0.001	3.700σ	1.327σ	145.839	18.759	2.788
0.0027	3.700σ	1.240σ	145.839	16.380	2.984
0.010	3.750σ	1.109σ	149.807	13.102	3.381
0.025	3.840σ	1.009σ	157.084	10.846	3.806
0.050	3.980σ	0.930σ	168.747	9.214	4.280
0.100	4.320σ	0.830σ	198.810	7.339	5.205

Table 9. Relationship between the MDB and MIB for Δh_4 and for $\tilde{\mathcal{P}}_{CD} = \tilde{\mathcal{P}}_{CI} = 80\%$.

α'	MIB	MDB	$\lambda_{q=1}^{(MIB)}$	$\lambda_{q=1}^{(MDB)}$	MIB/MDB
0.001	2.558σ	1.170σ	88.735	18.564	2.186
0.0027	2.566σ	1.093σ	89.291	16.201	2.348
0.010	2.598σ	0.982σ	91.532	13.077	2.646
0.025	2.659σ	0.895σ	95.902	10.863	2.971
0.050	2.784σ	0.820σ	105.107	9.118	3.395
0.100	3.082σ	0.738σ	128.771	7.390	4.174

Table 10. Relationship between the MDB and MIB for Δh_5 and for $\tilde{\mathcal{P}}_{CD} = \tilde{\mathcal{P}}_{CI} = 80\%$.

α'	MIB	MDB	$\lambda_{q=1}^{(MIB)}$	$\lambda_{q=1}^{(MDB)}$	MIB/MDB
0.001	11.290σ	3.065σ	252.065	18.577	3.684
0.0027	11.260σ	2.863σ	250.727	16.209	3.933
0.010	11.315σ	2.565σ	253.183	13.011	4.411
0.025	11.360σ	2.328σ	255.201	10.717	4.880
0.050	11.530σ	2.127σ	262.896	8.947	5.421
0.100	11.940σ	1.906σ	281.925	7.184	6.264

Table 11. Relationship between the MDB and MIB for Δh_6 and for $\tilde{\mathcal{P}}_{CD} = \tilde{\mathcal{P}}_{CI} = 80\%$.

α'	MIB	MDB	$\lambda_{q=1}^{(MIB)}$	$\lambda_{q=1}^{(MDB)}$	MIB/MDB
0.001	5.680σ	2.289σ	113.183	18.375	2.482
0.0027	5.700σ	2.134σ	113.981	15.976	2.671
0.010	5.695σ	1.908σ	113.781	12.769	2.985
0.025	5.825σ	1.729σ	119.035	10.492	3.368
0.050	6.021σ	1.579σ	127.180	8.747	3.813
0.100	6.394σ	1.409σ	143.426	6.965	4.538

The overall probabilities of wrong exclusion (\mathcal{P}_{WE}) for network (b) are provided in Figure 17. In general, we observe that the wrong exclusion rate (\mathcal{P}_{WE}) increases up to a certain outlier magnitude and, from this point on, the wrong exclusion rate (\mathcal{P}_{WE}) starts to decline, and the effect of the user-defined Type 1 decision error (α') on \mathcal{P}_{WE} becomes neutral in practical terms. This effect is due to the residuals' correlation. To see this effect more clearly, we also computed the individual contribution of each measurement to the overall wrong exclusion \mathcal{P}_{WE} and their corresponding weighting factors given by Equations (40) and (41), respectively.

The individual contributions to the overall \mathcal{P}_{WE} and their weighting factors for $\alpha' = 0.1$ are displayed in Figures 18 and 19, respectively. It is important to mention that the behaviour shown in Figures 18 and 19 is similar to that for other α' values. We observe that the correlation coefficient (ρ_{w_i, w_j}) only has a direct relationship with \mathcal{P}_{WE} for a certain outlier magnitude. Let us consider the

case in which Δh_6 is set up as an outlier. In that case, the larger the correlation coefficient (ρ_{w_i, w_j}), the higher the individual contribution to \mathcal{P}_{WE} . Of course, this only holds true if the outlier magnitude is larger than 3.2σ . This is also evident from the results of the weighting factors in Figure 19.

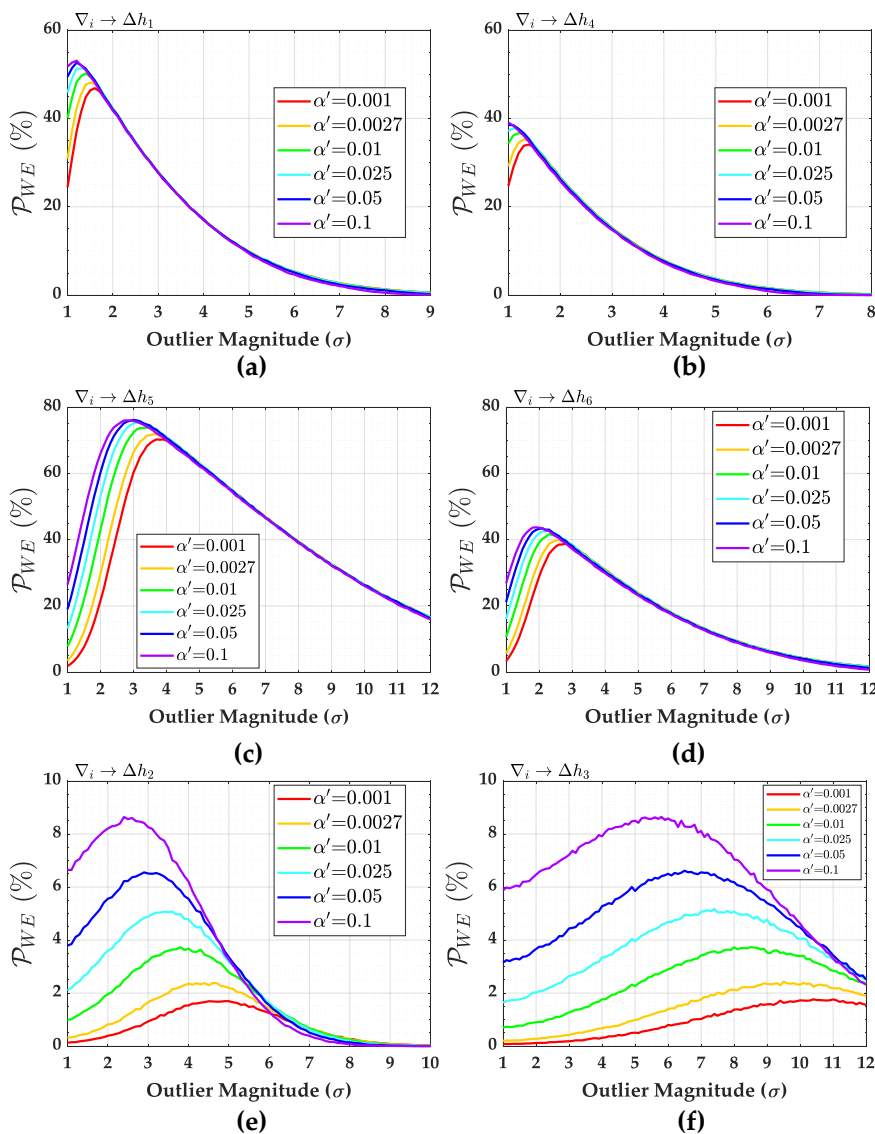


Figure 17. Probability of wrong exclusion (\mathcal{P}_{WE}) for network (b). (a) \mathcal{P}_{WE} for Δh_1 . (b) \mathcal{P}_{WE} for Δh_4 . (c) \mathcal{P}_{WE} for Δh_5 . (d) \mathcal{P}_{WE} for Δh_6 . (e) \mathcal{P}_{WE} for Δh_2 . (f) \mathcal{P}_{WE} for Δh_3 .

An important highlight is the association between the MIB and the contribution of each measurement to the probability of wrong exclusion \mathcal{P}_{WE} in Figure 18.

We observe that it is possible to find the value of the MIB at high success rates when the individual contributions to the overall wrong exclusion \mathcal{P}_{WE} of a given outlier start to decrease simultaneously. It is important to mention that this simultaneous decay occurs when there is a direct relationship between the correlation coefficient (ρ_{w_i, w_j}) and the wrong exclusion rate \mathcal{P}_{WE} . In that case, the identifiability of a given outlier can be verified for a given significance level α' and probability of correct identification \mathcal{P}_{CI} .

Figure 20 illustrates an example for measurements Δh_1 and Δh_4 . The black dashed line corresponds to the probability of correct identification \mathcal{P}_{CI} and the respective MIB for $\alpha' = 0.001$. Note that when the effect of all measurements on \mathcal{P}_{WE} decreases, it is possible to find an outlier

magnitude that can be identified. In other words, the effect of the correlation between residuals (ρ_{w_i,w_j}) becomes insignificant at a certain outlier magnitude, which increases the probability of identification.

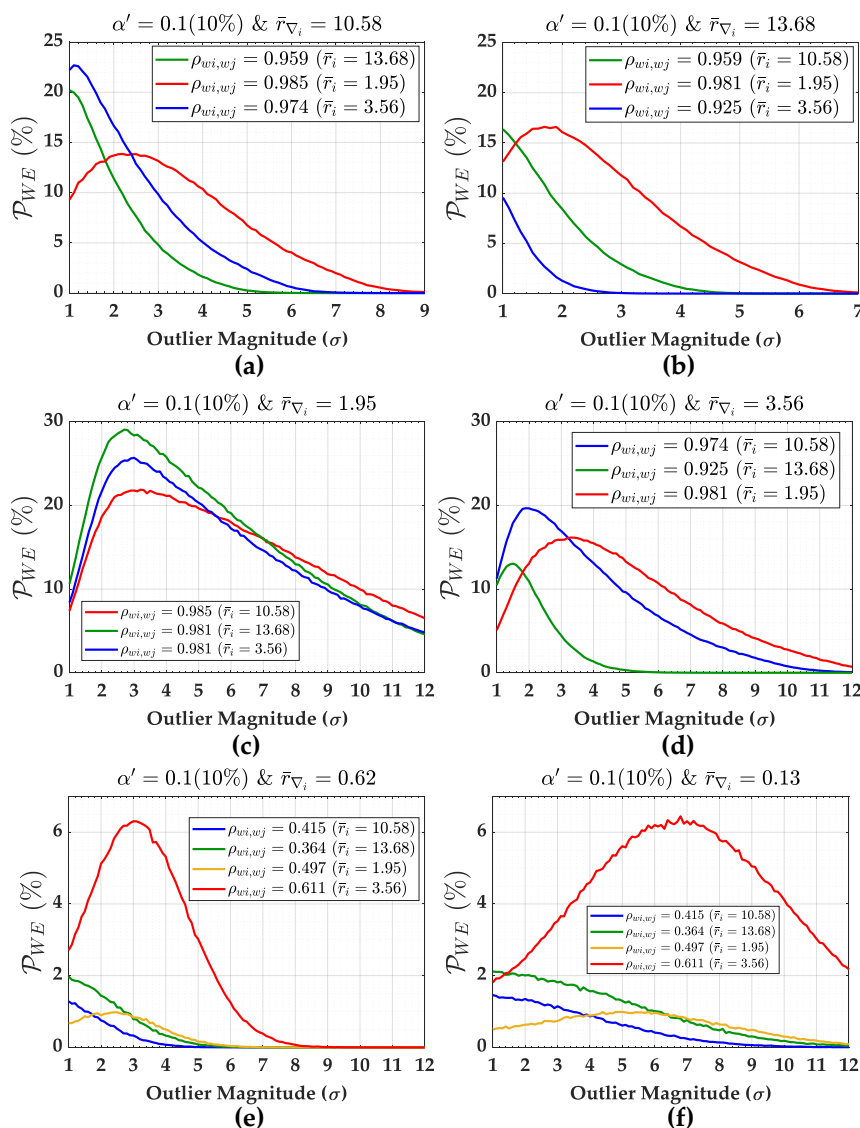


Figure 18. Individual contribution of each measurement to the overall wrong exclusion probability (\mathcal{P}_{WE}) for network (b) and for $\alpha' = 0.1$. (a) Individual contribution to \mathcal{P}_{WE} for Δh_1 . (b) Individual contribution to \mathcal{P}_{WE} for Δh_4 . (c) Individual contribution to \mathcal{P}_{WE} for Δh_5 . (d) Individual contribution to \mathcal{P}_{WE} for Δh_6 . (e) Individual contribution to \mathcal{P}_{WE} for Δh_2 . (f) Individual contribution to \mathcal{P}_{WE} for Δh_3 .

The probabilities of wrong exclusion for both Δh_2 and Δh_3 are smaller than those for the other cases. This is because of the correlation between residuals (ρ_{w_i,w_j}). In fact, we also note that although there is no reliability in terms of outlier identification for cases in which the correlation is $\rho_{w_i,w_j} = 1.00$ (i.e., 100%), there is reliability for outlier detection. In this case, outlier detection is caused by overlapping w -test statistics. The result for statistical overlap (\mathcal{P}_{ol}) is displayed in Figure 21. In general, the larger the Type 1 decision error α' , the larger the statistical overlap (\mathcal{P}_{ol}).

The over-identification cases (\mathcal{P}_{over+} and \mathcal{P}_{over-}) are displayed in Figures 22 and 23, respectively. We observe that the larger the Type I decision error (α'), the larger the over-identification cases. It should be noted that \mathcal{P}_{over+} is always larger than \mathcal{P}_{over-} . Over-identification \mathcal{P}_{over-} is practically null. The over-identification cases \mathcal{P}_{over+} for Δh_2 , Δh_3 and Δh_6 and \mathcal{P}_{over-} for Δh_1 are exactly null. The over-identifications \mathcal{P}_{over-} for Δh_2 and Δh_3 are less than 0.2% and are therefore not shown here.

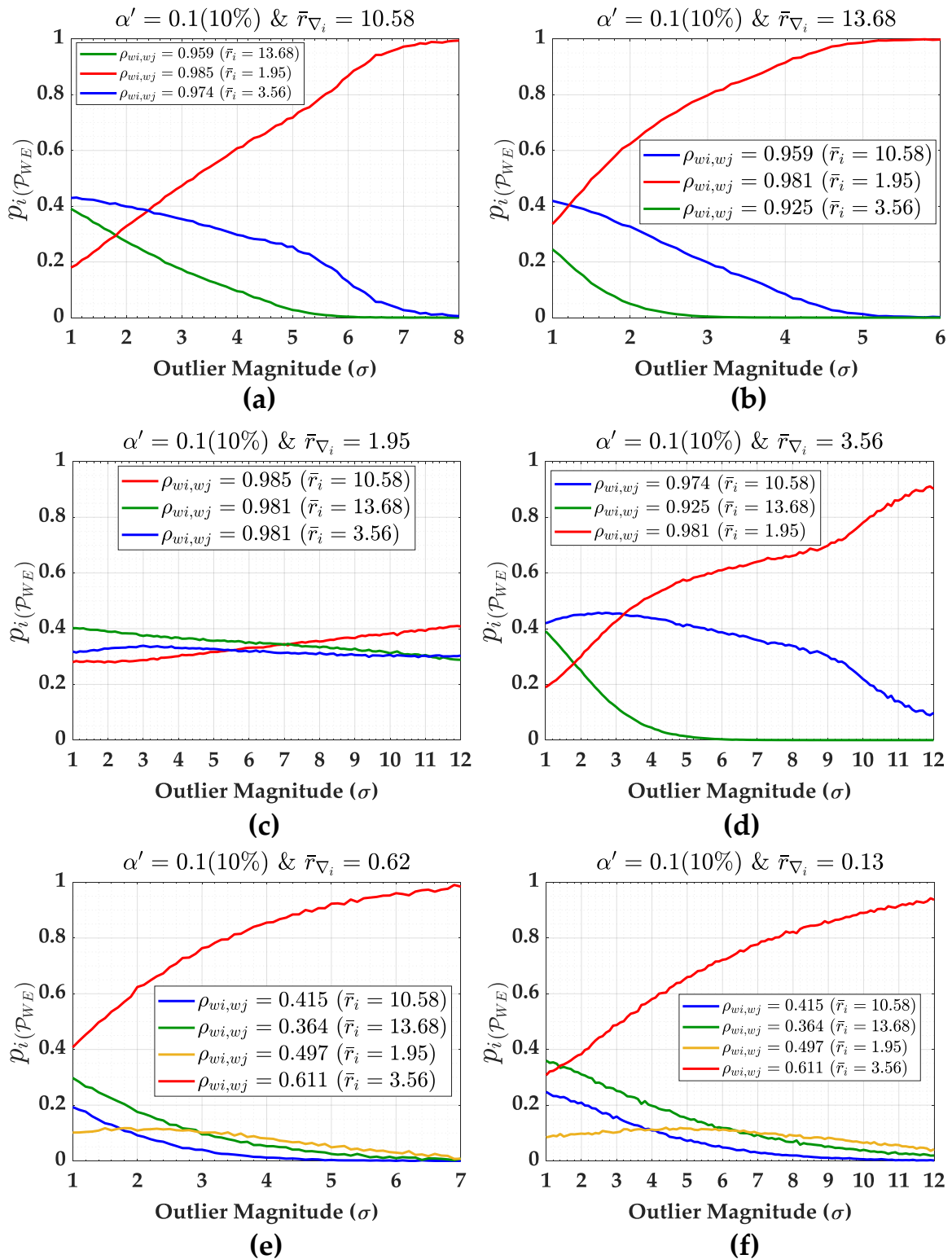


Figure 19. Weighting factors ($p_i(\mathcal{P}_{WE})$) of each measurement’s contribution to the overall wrong exclusion probability (\mathcal{P}_{WE}) for network (b) and for $\alpha' = 0.1$. (a) Weighting factors of contributions to \mathcal{P}_{WE} for Δh_1 . (b) Weighting factors of contributions to \mathcal{P}_{WE} for Δh_4 . (c) Weighting factors of contributions to \mathcal{P}_{WE} for Δh_5 . (d) Weighting factors of contributions to \mathcal{P}_{WE} for Δh_6 . (e) Weighting factors of contributions to \mathcal{P}_{WE} for Δh_2 . (f) Weighting factors of contributions to \mathcal{P}_{WE} for Δh_3 .

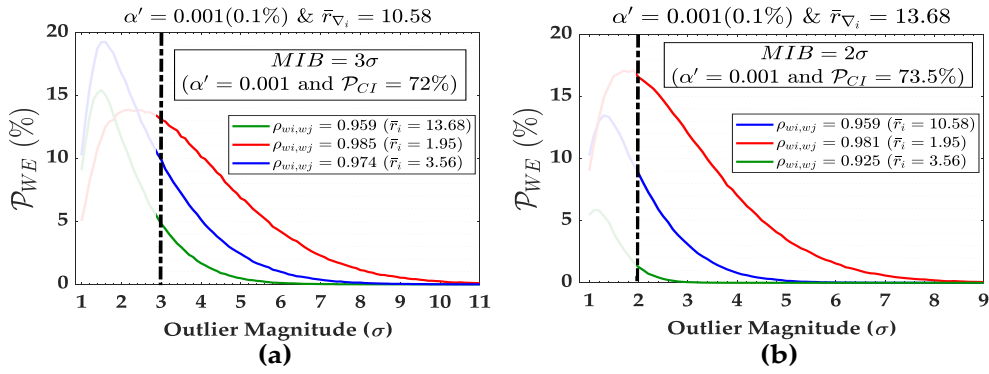


Figure 20. Relationship of the individual contributions to the overall probability of wrong exclusion (\mathcal{P}_{WE}) with correct identification rate \mathcal{P}_{CI} and MIB for $\alpha' = 0.001$. (a) Example for measurement Δh_4 . (b) Example for measurement Δh_1 .

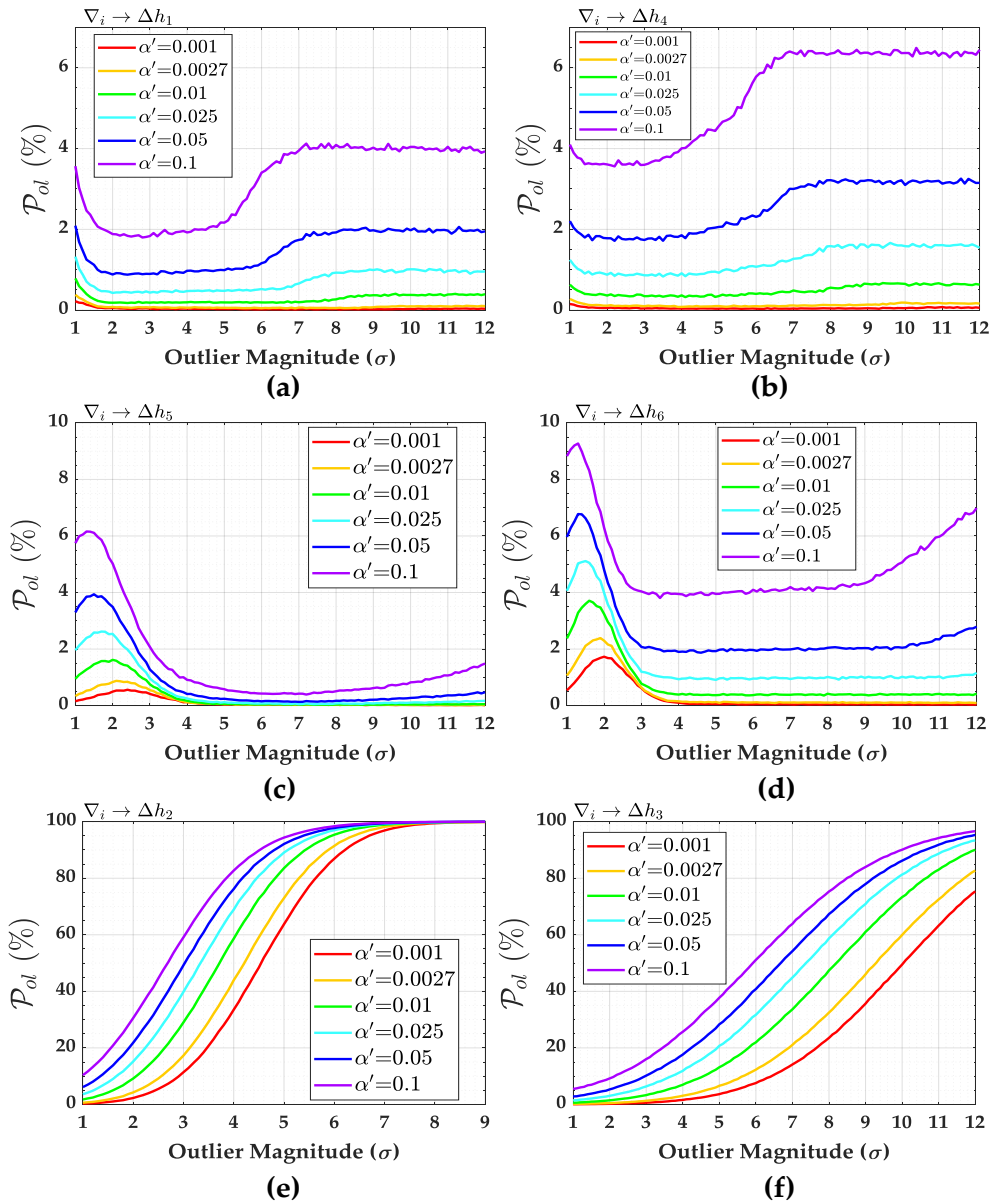


Figure 21. Probability of statistical overlap (\mathcal{P}_{ol}) for network (b). (a) \mathcal{P}_{ol} for Δh_1 . (b) \mathcal{P}_{ol} for Δh_4 . (c) \mathcal{P}_{ol} for Δh_5 . (d) \mathcal{P}_{ol} for Δh_6 . (e) \mathcal{P}_{ol} for Δh_2 . (f) \mathcal{P}_{ol} for Δh_3 .

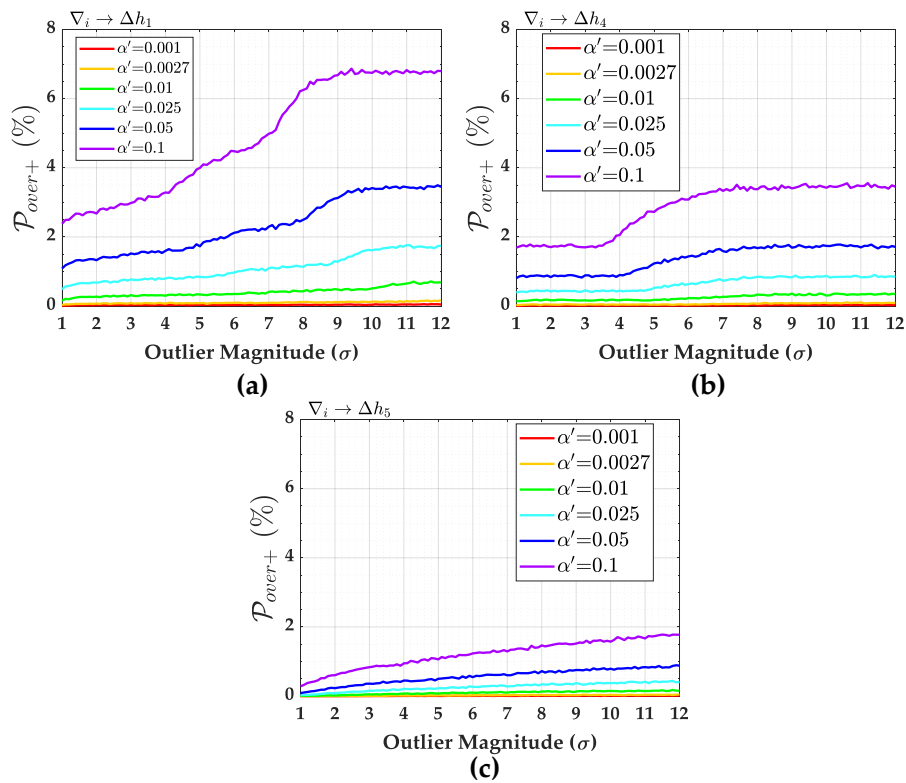


Figure 22. Probability of over-identification (\mathcal{P}_{over+}) for network (b). (a) \mathcal{P}_{over+} for Δh_1 . (b) \mathcal{P}_{over+} for Δh_4 . (c) \mathcal{P}_{over+} for Δh_5 .

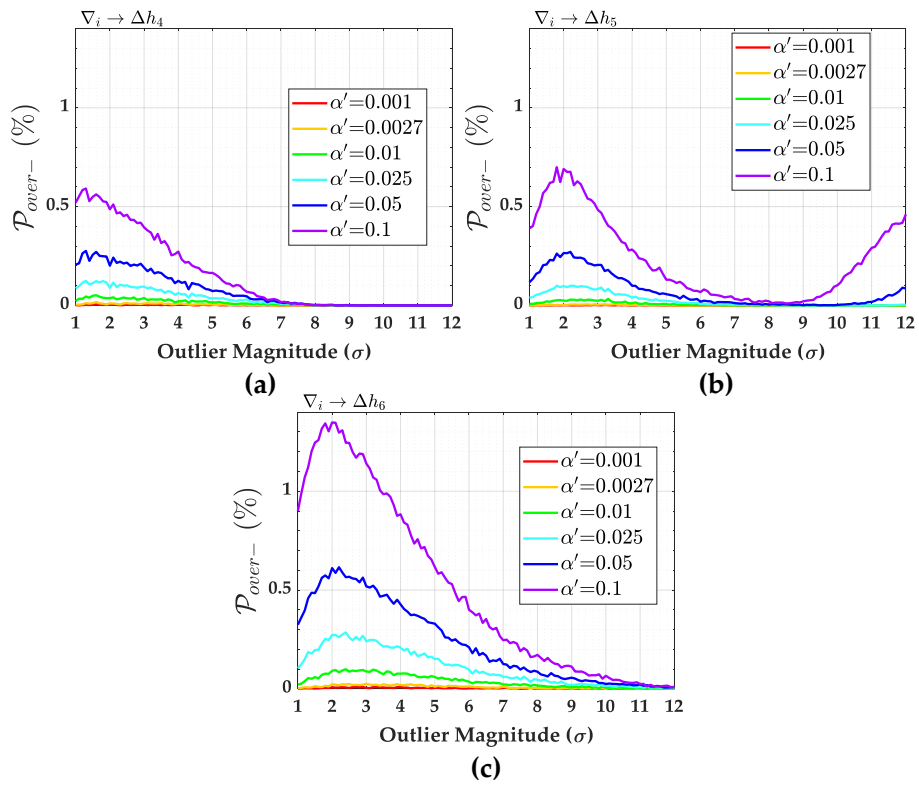


Figure 23. Probability of over-identification (\mathcal{P}_{over-}) for network (b). (a) \mathcal{P}_{over-} for Δh_4 . (b) \mathcal{P}_{over-} for Δh_5 . (c) \mathcal{P}_{over-} for Δh_6 .

7. Conclusions and Outlook

In this paper, we propose a procedure to compute the probability levels associated with an iterative outlier elimination procedure. This iterative outlier elimination procedure is known among geodesists as iterative data snooping *IDS*. On the basis of the probability levels of *IDS*, the sensitivity indicators—the Minimal Detectable Bias (MDB) and Minimal Identifiable Bias (MIB)—can also be determined for a given measurement system.

We emphasize that the probability levels associated with *IDS* in the presence of an outlier were analysed as a function of the user-defined Type I decision error (α'), outlier magnitude (∇_i), correlation between test statistics (ρ_{w_i, w_j}) and reliability indicators (i.e., redundancy number r_i and reliability number \bar{r}_i). It is important to highlight that these probability levels are based on critical values that were optimized via Monte Carlo.

We highlight the main findings of the paper below:

1. If one adopts the Bonferroni correction to compute the critical value of the test statistic associated with *IDS*, one does not have control over Type I decision errors. This is only true for small α' values and for a measurement system with high redundancy and a low correlation between test statistics.
2. If one maintains the condition of a measurement system with a low correlation between test statistics, the probability of wrong exclusion \mathcal{P}_{WE} is too low. In that case, one should opt for a larger α' so that the probability of missed detection \mathcal{P}_{MD} is as small as possible. Thus, it is possible to guarantee a high outlier identification rate. However, we verify that, under certain circumstances, the larger the Type I decision error α' , the higher the probability of correct detection \mathcal{P}_{CD} but the lower the probability of correct identification \mathcal{P}_{CI} . In that case, the larger the Type I Error α' , the larger the ratio between the sensitivity indicators MIB/MDB.
3. The larger the Type I error (α'), the higher the probability of correct outlier identification (\mathcal{P}_{CI}). However, it is valid only to a certain limit of outlier magnitude (threshold). There is an inversion when the outlier magnitude is greater than this threshold: i.e., the larger the α' , the lower the \mathcal{P}_{CI} . This is more critical in the case of a measurement system with a high correlation between test statistics. Moreover, the Type I decision error α' restricts the maximum rate of \mathcal{P}_{CI} .
4. We also observe that it is possible to find the value of the MIB when the contributions of each measurement to the probability of wrong exclusion \mathcal{P}_{WE} start to decline simultaneously. In that case, the identifiability of a given outlier can be verified for a given α' and \mathcal{P}_{CI} . In other words, for a certain outlier magnitude, the effect of the correlation between test statistics becomes insignificant, which increases the probability of identification. Moreover, if a small outlier magnitude (outlier with a magnitude close to the measurement uncertainty) were to arise for a measurement system with a high correlation between test statistics, the alternative hypotheses would not be distinguished; i.e., this outlier would never be identified.
5. The larger the Type I decision error α' , the larger the over-identification cases. The over-identification case \mathcal{P}_{over+} is always larger than \mathcal{P}_{over-} . We also note that the lower the correlation between test statistics, the higher the probability of over-identification positive \mathcal{P}_{over+} . For small α' (close to $\alpha' = 0.001$), \mathcal{P}_{over-} is practically null.
6. When the correlation between two test statistics is equal to exactly 1.00, \mathcal{P}_{CI} does not exist, but there is \mathcal{P}_{CD} , which is mainly caused by \mathcal{P}_{ol} .

The computation procedure presented in this paper was successfully applied to a practical example of geodetic networks. Although the procedure was applied to geodetic networks, it is a generally applicable method. The authors have been working on solutions to find a relationship between the variables computed deterministically (e.g., local redundancy, residuals' correlation) with the probability levels computed by Monte Carlo. The use of Monte Carlo will no longer be needed to find the MIB if a model is found. Moreover, further investigation is required to apply this analysis to general problems with multiple outliers.

Author Contributions: Conceptualization, V.F.R., M.T.M. and I.K.; methodology, V.F.R.; software, V.F.R. And L.G.d.S.J.; validation, V.F.R., M.T.M. and I.K.; formal analysis, V.F.R., M.T.M. and I.K.; investigation, V.F.R.; data curation, V.F.R., I.K. and M.T.M.; writing—original draft preparation, V.F.R.; writing—review and editing, V.F.R., M.T.M., I.K. and M.R.V.; supervision, M.T.M., I.K., M.R.V. and L.G.d.S.J.; project administration, M.T.M. and I.K.; funding acquisition, M.T.M., M.R.V. and L.G.d.S.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the CNPq—Conselho Nacional de Desenvolvimento Científico e Tecnológico—Brasil (proc. n^o 103587/2019-5). This research and the APC were also funded by PETROBRAS (Grant Number 2018/00545-0).

Acknowledgments: The authors would like to thank the three anonymous reviewers who contributed to the improvement of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, J.; Knight, N.L. New Outlier Separability Test and Its Application in GNSS Positioning. *J. Glob. Position. Syst.* **2012**, *11*, 46–57. [[CrossRef](#)]
2. Yang, L.; Wang, J.; Knight, N.L.; Shen, Y. Outlier separability analysis with a multiple alternative hypotheses test. *J. Geod.* **2013**, *87*, 591–604. [[CrossRef](#)]
3. Teunissen, P.J.G.; Imperato, D.; Tiberius, C.C.J.M. Does RAIM with Correct Exclusion Produce Unbiased Positions? *Sensors* **2017**, *17*, 1508. [[CrossRef](#)] [[PubMed](#)]
4. Na, W.; Park, C.; Lee, S.; Yu, S.; Lee, H. Sensitivity-Based Fault Detection and Isolation Algorithm for Road Vehicle Chassis Sensors. *Sensors* **2018**, *18*, 2720. [[CrossRef](#)] [[PubMed](#)]
5. Crispoltoni, M.; Fravolini, M.L.; Balzano, F.; D’Urso, S.; Napolitano, M.R. Interval Fuzzy Model for Robust Aircraft IMU Sensors Fault Detection. *Sensors* **2018**, *18*, 2488. [[CrossRef](#)]
6. Nguyen, V.K.; Renault, E.; Milocco, R. Environment Monitoring for Anomaly Detection System Using Smartphones. *Sensors* **2019**, *19*, 3834. [[CrossRef](#)]
7. Mei, X.; Wu, H.; Xian, J.; Chen, B.; Zhang, H.; Liu, X. A Robust, Non-Cooperative Localization Algorithm in the Presence of Outlier Measurements in Ocean Sensor Networks. *Sensors* **2019**, *19*, 2708. [[CrossRef](#)]
8. Nie, Y.; Yang, L.; Shen, Y. Specific Direction-Based Outlier Detection Approach for GNSS Vector Networks. *Sensors* **2019**, *19*, 1836. [[CrossRef](#)]
9. Leslar, M.; Wang, J.G.; Hu, B. Comprehensive Utilization of Temporal and Spatial Domain Outlier Detection Methods for Mobile Terrestrial LiDAR Data. *Remote. Sens.* **2011**, *3*, 1724–1742. [[CrossRef](#)]
10. Rofatto, V.F.; Matsuoka, M.T.; Klein, I.; Veronez, M.R. Monte-Carlo-based uncertainty propagation in the context of Gauss–Markov model: a case study in coordinate transformation. *Sci. Plena* **2019**, *15*, 1–17. [[CrossRef](#)]
11. Lehmann, R. Observation error model selection by information criteria vs. normality testing. *Stud. Geophys. Geod.* **2015**, *59*, 489–504. [[CrossRef](#)]
12. 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. *Surv. Rev.* **2018**, 1–17. [[CrossRef](#)]
13. Lehmann, R. On the formulation of the alternative hypothesis for geodetic outlier detection. *J. Geod.* **2013**, *87*, 373–386. [[CrossRef](#)]
14. Rousseeuw, P.J.; Leroy, A.M. *Robust Regression and Outlier Detection*, 1st ed.; Wiley-Interscience: Hoboken, NJ, USA, 2003.
15. Yang, Y. Robust estimation of geodetic datum transformation. *J. Geod.* **1999**, *73*, 268–274. [[CrossRef](#)]
16. Wilcox, R. *Introduction to Robust Estimation and Hypothesis Testing*, 3rd ed.; Academic Press: Cambridge, MA, USA, 2013. [[CrossRef](#)]
17. Duchnowski, R. Hodges–Lehmann estimates in deformation analyses. *J. Geod.* **2013**, *87*, 873–884. [[CrossRef](#)]
18. Klein, I.; Matsuoka, M.T.; Guzzato, M.P.; de Souza, S.F.; Veronez, M.R. On evaluation of different methods for quality control of correlated observations. *Surv. Rev.* **2015**, *47*, 28–35. [[CrossRef](#)]
19. Baarda, W. Statistical Concepts in Geodesy. *Publ. Geod. Neth. Geod. Comm.* **1967**, *2*, 1–74.
20. Baarda, W. A Testing Procedure for Use in Geodetic Networks. *Publ. Geod. Neth. Geod. Comm.* **1968**, *2*, 1–97.

21. Förstner, W. Reliability and discernability of extended Gauss-Markov models. In *Seminar on Mathematical Models to Outliers and Systematic Errors*; No. 98; German Geodetic Commission (DGK): Munich, Germany, 1983; Volume A, pp. 79–103.
22. Lehmann, R. Improved critical values for extreme normalized and studentized residuals in Gauss–Markov models. *J. Geod.* **2012**, *86*, 1137–1146. [[CrossRef](#)]
23. Prószyński, W. Revisiting Baarda’s concept of minimal detectable bias with regard to outlier identifiability. *J. Geod.* **2015**, *89*, 993–1003. [[CrossRef](#)]
24. Marshall, J. L1-norm pre-analysis measures for geodetic networks. *J. Geod.* **2002**, *76*, 334–344. [[CrossRef](#)]
25. Huber, P.J. Robust Statistics. In *International Encyclopedia of Statistical Science*; Lovric, M., Ed.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 1248–1251. [[CrossRef](#)]
26. WiÅniewski, Z. Estimation of parameters in a split functional model of geodetic observations (M split estimation). *J. Geod.* **2008**, *82*. [[CrossRef](#)]
27. WiÅniewski, Z.; Duchnowski, R.; Dumalski, A. Efficacy of Msplit Estimation in Displacement Analysis. *Sensors* **2019**, *19*, 47. [[CrossRef](#)] [[PubMed](#)]
28. Hodges, J.L.; Lehmann, E.L. Estimates of Location Based on Rank Tests. *Ann. Math. Stat.* **1963**, *34*, 598–611. [[CrossRef](#)]
29. Duchnowski, R. Robustness of Strategy for Testing Levelling Mark Stability Based on Rank Tests. *Surv. Rev.* **2011**, *43*, 687–699. [[CrossRef](#)]
30. Wyszowska, P.; Duchnowski, R. Subjective breakdown points of R-estimators applied in deformation analysis. In Proceedings of the International Conference on Environmental Engineering, Vilnius, Lithuania, 27–28 April 2017; pp. 1–6. [[CrossRef](#)]
31. Koch, I.É.; Klein, I.; Gonzaga, L.; Matsuoka, M.T.; Rofatto, V.F.; Veronez, M.R. Robust Estimators in Geodetic Networks Based on a New Metaheuristic: Independent Vortices Search. *Sensors* **2019**, *19*, 4535. [[CrossRef](#)]
32. Lehmann, R. 3σ -Rule for Outlier Detection from the Viewpoint of Geodetic Adjustment. *J. Surv. Eng.* **2013**, *139*, 157–165. [[CrossRef](#)]
33. Lehmann, R.; Scheffler, T. Monte Carlo based data snooping with application to a geodetic network. *J. Appl. Geod.* **2011**, *5*, 123–134. [[CrossRef](#)]
34. Koch, K.R. *Parameter Estimation and Hypothesis Testing in Linear Models*, 2nd ed.; Springer: Berlin, Germany, 1999.
35. Teunissen, P. *Testing Theory: An Introduction*, 2nd ed.; Delft University Press: Delft, The Netherlands, 2006.
36. Ghilani, C.D. *Adjustment Computations: Spatial Data Analysis*, 6th ed.; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2017.
37. Zaminpardaz, S.; Teunissen, P. DIA-datasnooping and identifiability. *J. Geod.* **2019**, *93*, 85–101. [[CrossRef](#)]
38. Zhao, Y.; Sun, R.; Ni, Z. Identification of Natural and Anthropogenic Drivers of Vegetation Change in the Beijing-Tianjin-Hebei Megacity Region. *Remote Sens.* **2019**, *11*, 1224. [[CrossRef](#)]
39. Wang, K.N.; Ao, C.O.; Juárez, M.D. GNSS-RO Refractivity Bias Correction Under Ducting Layer Using Surface-Reflection Signal. *Remote Sens.* **2020**, *12*, 359. [[CrossRef](#)]
40. Lee, G. An Efficient Compressive Hyperspectral Imaging Algorithm Based on Sequential Computations of Alternating Least Squares. *Remote Sens.* **2019**, *11*, 2932. [[CrossRef](#)]
41. Zhang, Y.; Wang, X.; Balzter, H.; Qiu, B.; Cheng, J. Directional and Zonal Analysis of Urban Thermal Environmental Change in Fuzhou as an Indicator of Urban Landscape Transformation. *Remote Sens.* **2019**, *11*, 2810. [[CrossRef](#)]
42. Kok, J.J.; States, U. *On Data Snooping and Multiple Outlier Testing*, NOAA Technical Report NOS, NGS, 30; U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Charting and Geodetic Services: Rockville, MD, USA, 1984.
43. Knight, N.L.; Wang, J.; Rizos, C. Generalised measures of reliability for multiple outliers. *J. Geod.* **2010**, *84*, 625–635. [[CrossRef](#)]
44. Gui, Q.; Li, X.; Gong, Y.; Li, B.; Li, G. A Bayesian unmasking method for locating multiple gross errors based on posterior probabilities of classification variables. *J. Geod.* **2011**, *85*, 191–203. [[CrossRef](#)]
45. Klein, I.; Matsuoka, M.T.; Guzzato, M.P.; Nievinski, F.G. An approach to identify multiple outliers based on sequential likelihood ratio tests. *Surv. Rev.* **2017**, *49*, 449–457. [[CrossRef](#)]
46. Imparato, D.; Teunissen, P.; Tiberius, C. Minimal Detectable and Identifiable Biases for quality control. *Surv. Rev.* **2019**, *51*, 289–299. [[CrossRef](#)]

47. Teunissen, P.J.G. Distributional theory for the DIA method. *J. Geod.* **2018**, *92*, 59–80. [[CrossRef](#)]
48. Hekimoglu, S.; Koch, K.R. How can reliability of the robust methods be measured? In *Third Turkish-German Joint Geodetic Days: Towards a Digital Age*; Altan, M.O., Gründig, L., Eds.; Istanbul Technical University: Istanbul, Turkey, 1999; Volume 1, pp. 179–196.
49. Aydin, C. Power of Global Test in Deformation Analysis. *J. Surv. Eng.* **2012**, *138*, 51–56. [[CrossRef](#)]
50. Nowel, K. Application of Monte Carlo method to statistical testing in deformation analysis based on robust M-estimation. *Surv. Rev.* **2016**, *48*, 212–223. [[CrossRef](#)]
51. Klein, I.; Matsuoka, M.T.; Guzzato, M.P.; Nievinski, F.G.; Veronez, M.R.; Rofatto, V.F. A new relationship between the quality criteria for geodetic networks. *J. Geod.* **2019**, *93*, 529–544. [[CrossRef](#)]
52. Robert, C.; Casella, G. *Monte Carlo Statistical Methods*, 2 ed.; Springer: New York, NY, USA, 2004.
53. Gamerman, D.; Lopes, H. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd ed.; Taylor & Francis: London, UK, 2006.
54. Koch, K. Bayesian statistics and Monte Carlo methods. *J. Geod. Sci.* **2018**, *8*, 18–29. [[CrossRef](#)]
55. Rofatto, V.; Matsuoka, M.; Klein, I. An Attempt to Analyse Baarda's Iterative Data Snooping Procedure based on Monte Carlo Simulation. *S. Afr. J. Geomat.* **2017**, *6*, 416–435. [[CrossRef](#)]
56. Bonferroni, C. Teoria Statistica Delle Classi E Calcolo Delle Probabilità. *Pubbl. d. R. Ist. Super. di Sci. Econom. e Commerciali di Firenze* **1936**, *8*, 1–62.
57. Velsink, H. On the deformation analysis of point fields. *J. Geod.* **2015**, *89*, 1071–1087. [[CrossRef](#)]
58. Lehmann, R.; Lösler, M. Multiple Outlier Detection: Hypothesis Tests versus Model Selection by Information Criteria. *J. Surv. Eng.* **2016**, *142*, 04016017. [[CrossRef](#)]
59. Lehmann, R.; Lösler, M. Congruence analysis of geodetic networks—Hypothesis tests versus model selection by information criteria. *J. Appl. Geod.* **2017**, *11*, 271–283. [[CrossRef](#)]
60. Rofatto, V.; Matsuoka, M.; Klein, I. Design of geodetic networks based on outlier identification criteria: An example applied to the leveling network. *Bull. Geod. Sci.* **2018**, *24*, 152–170. [[CrossRef](#)]
61. Matsuoka, M.T.; Rofatto, V.F.; Klein, I.; Roberto Veronez, M.; da Silveira, L.G.; Neto, J.B.S.; Alves, A.C.R. Control Points Selection Based on Maximum External Reliability for Designing Geodetic Networks. *Appl. Sci.* **2020**, *10*, 687. [[CrossRef](#)]
62. Koch, K.R. Expectation Maximization algorithm and its minimal detectable outliers. *Stud. Geophys. Geod.* **2017**, *61*, 1–18. [[CrossRef](#)]
63. Arnold, S. *The Theory of Linear Models and Multivariate Analysis*, 1st ed.; Wiley: Hoboken, NJ, USA, 1981.
64. Teunissen, P.J.G. An Integrity and Quality Control Procedure for Use in Multi Sensor Integration In Proceedings of the 3rd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1990), Colorado Spring, CO, USA, 19–21 September 1990; pp. 513–522.
65. Aydin, C.; Demirel, H. Computation of Baarda's lower bound of the non-centrality parameter. *J. Geod.* **2004**, *78*, 437–441. [[CrossRef](#)]
66. Mierlo, J.V. Statistical Analysis of Geodetic Measurements for the Investigation of Crustal Movements. In *Recent Crustal Movements, 1977*; Whitten, C., Green, R., Meade, B., Eds.; Elsevier: Amsterdam, The Netherlands, 1979; Volume 13, pp. 457–467. [[CrossRef](#)]
67. Hawkins, D.M. *Identification of Outliers*, 1st ed.; Springer: Amsterdam, The Netherlands, 1980. [[CrossRef](#)]
68. Van der Marel, H.; Rösters, A.J.M. Statistical Testing and Quality Analysis in 3-D Networks (part II) Application to GPS. In *Global Positioning System: An Overview*; Bock, Y., Leppard, N., Eds.; Springer: New York, NY, USA, 1990; pp. 290–297.
69. Romano, J.P.; Wolf, M. Multiple Testing of One-Sided Hypotheses: Combining Bonferroni and the Bootstrap. In *Predictive Econometrics and Big Data*; Kreinovich, V., Sriboonchitta, S., Chakpitak, N., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 78–94.
70. Bonimani, M.; Rofatto, V.; Matsuoka, M.; Klein, I. Application of artificial random numbers and Monte Carlo method in the reliability analysis of geodetic networks. *Rev. Bras. Comp. Apl.* **2019**, *11*, 74–85. [[CrossRef](#)]
71. Altiok, T.; Melamed, B. *Simulation Modeling and Analysis with Arena*, 1st ed.; Academic Press: Cambridge, MA, USA, 2007.
72. Matsumoto, M.; Nishimura, T. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.* **1998**, *8*, 3–30. [[CrossRef](#)]
73. Box, G.E.P.; Muller, M.E. A Note on the Generation of Random Normal Deviates. *Ann. Math. Stat.* **1958**, *29*, 610–611. [[CrossRef](#)]

74. Lemeshko, B.Y.; Lemeshko, S.B. Extending the Application of Grubbs-Type Tests in Rejecting Anomalous Measurements. *Meas. Tech.* **2005**, *48*, 536–547. [[CrossRef](#)]
75. Algarni, D.A.; Ali, A.E. Heighting and Distance Accuracy with Electronic Digital Levels. *J. King Saud Univ. Eng. Sci.* **1998**, *10*, 229–239. [[CrossRef](#)]
76. Gemin, A.R.; Matos, É.S.; Faggion, P.L. Application of calibration certificate of digital leveling systems in the monitoring of structures: A case study at the governador josã richa hydroelectric power plant-pr. *Boletim CiÃ GeodÃ* **2018**, *24*, 235–249. [[CrossRef](#)]
77. Takalo, M.; Rouhiainen, P. Development of a System Calibration Comparator for Digital Levels in Finland. *Nord. J. Surv. Real Estate Res.* **2004**, *1*, 119–130.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

4.5 Avaliação dos efeitos das injunções absolutas e relativas sobre a sensibilidade do Data-Snooping

É muito comum construir modelos matemáticos com base em algum conhecimento inicial sobre um determinado problema. Em outras palavras, os modelos geralmente são configurados de uma maneira que seus parâmetros atendam certas restrições. Tais restrições se originam de um conhecimento prévio do cientista. Dentro do contexto de geodésia, por exemplo, é comum ter um certo conhecimento prévio sobre a estabilidade de certos pontos em aplicações de monitoramento de estruturas. Tais informações são incorporadas no modelo matemático com o objetivo de auxiliar a detecção e identificação de possíveis deformações na estrutura monitorada. De modo geral, as restrições são usadas para, por exemplo:

- garantir a estabilidade das estimativas dos parâmetros no modelo;
- melhorar a acurácia dos resultados;
- reduzir o número de parâmetros desnecessários (insignificantes para o modelo);
- aumentar a redundância do sistema;
- auxiliar na detecção e identificação de possíveis outliers nos dados observados;
- auxiliar na parametrização de efeitos sistemáticos.

Os modelos podem ser formulados com restrições mínimas ou restrições redundantes. Na maioria das vezes, restrições mínimas são introduzidas para resolver o problema de deficiência de característica (posto) em sistemas lineares (ou linearizados). A deficiência de posto geralmente é causada pela falta (ou insuficiência) de informações sobre um problema. No campo da geodésia, por exemplo, as restrições mínimas têm a função principal de especificar o sistema de coordenadas no qual as posições de pontos, levantados por medidas de direção e distâncias, serão determinadas por algum método de estimação (Mínimos quadrados, BLUE e/ou MLE). Nesse contexto, Datum e Sistema de Referência, por exemplo, são conceitos comuns usados em geodésia e, por muitas vezes, são confundidos. O nosso objetivo aqui não é adentrar no mérito das definições, mas podemos antecipar que Datum são restrições matemáticas necessárias para a modelagem da forma e da dimensão da Terra, bem como do campo de gravidade e de suas variações temporais.

As restrições redundantes são informações que extrapolam o mínimo necessário para estimar os parâmetros do modelo. Essas restrições são normalmente usadas para verificar a estabilidade de pontos em análise de deformações geodésicas ou para testar a compatibilidade entre as observações e as restrições (VELSINK, 2015; LEHMANN; NEITZEL, 2013).

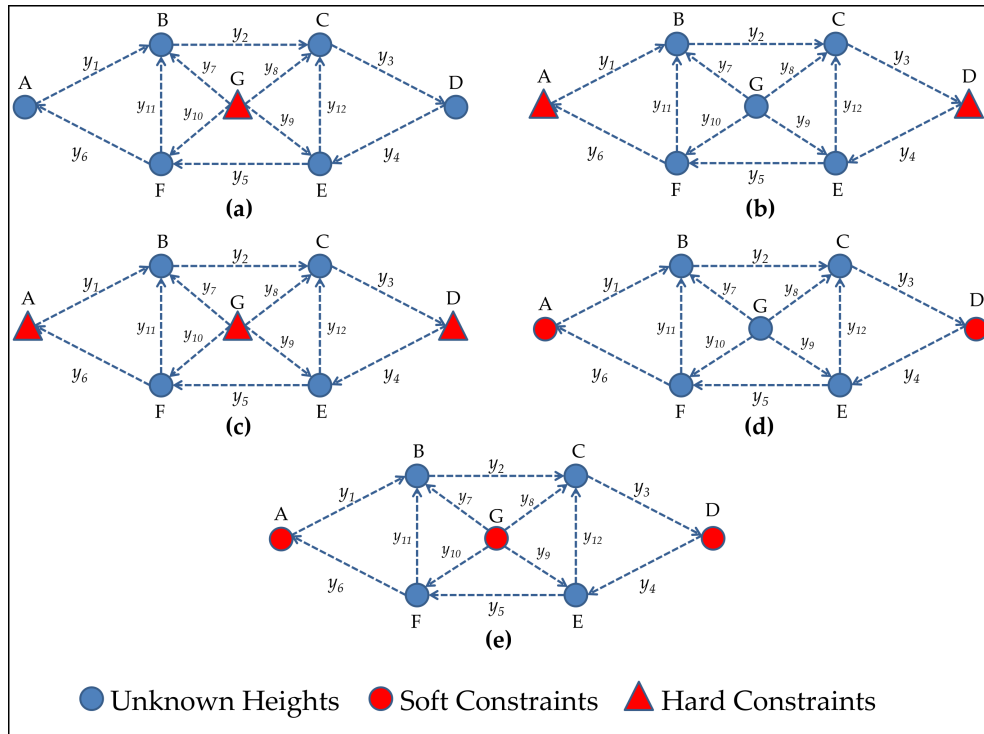


Figura 13 – Configurações da rede geodésica sujeita às diferentes injunções: altitude desconhecida (unknown heights) representado por círculo na cor azul, injunção absoluta (hard constraints) por triângulo em vermelho e injunção relativa (soft constraints) por círculo em vermelho.

As restrições também podem ser classificadas como rígidas (*Hard Constraints*), conhecidas em geodésia como injunção absoluta, e flexíveis (*Soft Constraints*), conhecidas em geodésia como injunção relativa. As injunções absolutas geralmente representam um caso em que existem relações funcionais exatamente conhecidas entre os parâmetros desconhecidos. Nesse caso, as injunções são tratadas como constantes e não como variáveis aleatórias. Por outro lado, as injunções relativas são tratadas como variáveis aleatórias e, portanto, podem violar tais relações funcionais em função de suas incertezas (AMIRI-SIMKOOEI, 2017; AMIRI-SIMKOOEI, 2007). Esse tipo de restrição também é conhecida como pseudo-observação (*pseudo-observation model*), pois é tratada simultaneamente como observação e parâmetro no processo de estimação.

Aqui, nós analisamos os efeitos das injunções mínimas, redundantes, absolutas e relativas sobre os níveis de probabilidade do DSI. Uma rede geodésica com doze observações não correlacionadas foi selecionada para realizar os experimentos. As configurações das injunções são mostradas na Figura (13).

No total, onze configurações de injunções foram consideradas, como segue:

- Rede minimamente injuncionada com uma restrição mínima. É necessário pelo menos um ponto de altitude conhecida para estimar as altitudes desconhecidas dos outros

pontos. Nesse caso, a redundância global do sistema foi $r = n - \text{car}(A) = n - u = 12 - 6 = 6$.

- Rede com duas injunções absolutas, ou seja, rede com restrições redundantes. Neste caso, a redundância global do sistema foi $r = 7$.
- Rede com três injunções absolutas com $r = 8$.
- Rede com duas injunções relativas com $r = 7$.
- Rede com três injunções relativas com $r = 8$.

No cenário das injunções relativas, o modelo estocástico foi configurado da seguinte forma: precisão das restrições dez vezes pior do que das observações (i.e. $\sigma_c = 10 \times \sigma = 10mm$); precisão dez vezes melhor (i.e. $\sigma_c = 0.1mm$) e restrições com grau de precisão igual ao das observações (i.e. $\sigma_c = 1mm$), em que σ_c é o desvio-padrão das restrições (*constraints*).

A Tabela (2) fornece a redundância local (r_i), desvio-padrão do outlier ($\sigma_{\hat{v}_i}$) e a máxima correlação absoluta entre os resíduos ($\max_{\rho_{w_i, w_j}}$) para cada cenário de restrição rígida.

Pela primeira em análises de redes geodésicas, uma técnica de *clustering* (agrupamento de dados) foi desenvolvida com o objetivo de reduzir a dimensão dos dados. As medidas com semelhante máxima correlação absoluta entre os resíduos (2.49) e redundância local (2.6) foram agrupadas. Para cada classe de probabilidade (\mathcal{P}_{CI} , \mathcal{P}_{CD} , \mathcal{P}_{WE} , \mathcal{P}_{over+} , \mathcal{P}_{over-} e \mathcal{P}_{ol}) e para cada grupo, o valor médio foi obtido, dando origem aos *clusters*. Para o caso de injunções absolutas, as doze observações $y_i, \forall i = 1, \dots, 12$ foram agrupadas em 4 clusters. As probabilidades de correta detecção (\mathcal{P}_{CD}) e identificação (\mathcal{P}_{CI}) para o caso de injunções absolutas, denotadas por **h.c.** (*hard constraints*), são mostradas na Figura (14). Os valores críticos foram calculados para $\alpha' = 0.001$, de acordo com o procedimento descrito no Capítulo 3, Seção 3.1.

Em termos de detecção e identificação de outlier, Cluster 1 foi mais sensível a restrições, o Cluster 3 relativamente sensível a restrições, enquanto o Cluster 4 foi completamente insensível a restrições, e o Cluster 2 relativamente insensível a restrições. A razão para isso é que a redundância local (r_i) do Cluster 1 elevou com o aumento do número de injunções absolutas, enquanto o Cluster 4 permaneceu o mesmo.

A Figura (15) mostra a probabilidade de falsa exclusão (\mathcal{P}_{WE}) para os casos de injunção absoluta.

Importante destacar que não existe confiabilidade em termos de identificação para o Cluster 1, como pode ser visto na Figura (14a). No caso do Cluster 1, a correta detecção

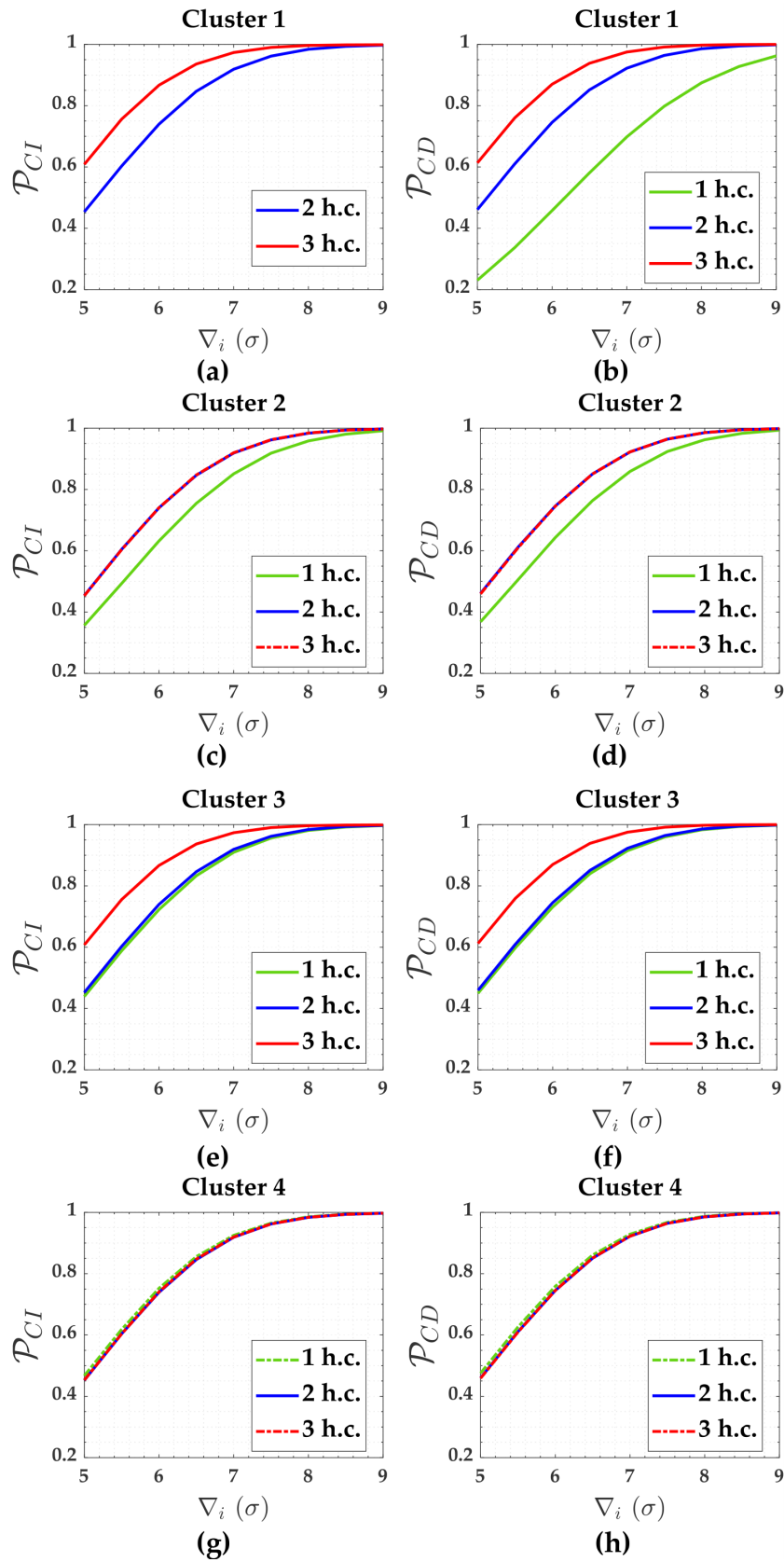


Figura 14 – Probabilidades de correta identificação (\mathcal{P}_{CI}) e detecção (\mathcal{P}_{CD}) para $\hat{k} = 3.89$, no caso de uma única injeção absoluta, e $\hat{k} = 3.93$ para o caso de duas e três injeções absoluta. Cluster 1 (a,b): y_1, y_3, y_4 e y_6 . Cluster 2 (c,d): y_2 e y_5 . Cluster 3 (e,f): y_7, y_8, y_9 e y_{10} . Cluster 4 (g,h): y_{11} e y_{12} .

Tabela 2 – Redundância local (r_i), desvio-padrão do outlier ($\sigma_{\hat{\nabla}_i}$) e máxima correlação absoluta entre os resíduos ($max_{\rho_{w_i, w_j}}$) para cada cenário de restrição rígida.

Observação	1 restrição rígida			2 injunções absolutas			3 injunções absolutas		
	r_i	$\sigma_{\hat{\nabla}_i}$	$max_{\rho_{w_i, w_j}}$	r_i	$\sigma_{\hat{\nabla}_i}$	$max_{\rho_{w_i, w_j}}$	r_i	$\sigma_{\hat{\nabla}_i}$	$max_{\rho_{w_i, w_j}}$
y_1	0.396	1.589	1.00	0.583	1.309	0.36	0.708	1.188	0.41
y_2	0.500	1.414	0.47	0.583	1.309	0.36	0.583	1.309	0.32
y_3	0.396	1.589	1.00	0.583	1.309	0.36	0.708	1.188	0.41
y_4	0.396	1.589	1.00	0.583	1.309	0.36	0.708	1.188	0.41
y_5	0.500	1.414	0.47	0.583	1.309	0.36	0.583	1.309	0.32
y_6	0.396	1.589	1.00	0.583	1.309	0.36	0.708	1.188	0.41
y_7	0.563	1.333	0.47	0.583	1.309	0.36	0.708	1.188	0.41
y_8	0.563	1.333	0.47	0.583	1.309	0.36	0.708	1.188	0.41
y_9	0.563	1.333	0.47	0.583	1.309	0.36	0.708	1.188	0.41
y_{10}	0.563	1.333	0.47	0.583	1.309	0.36	0.708	1.188	0.41
y_{11}	0.583	1.309	0.43	0.583	1.309	0.36	0.583	1.309	0.32
y_{12}	0.583	1.309	0.43	0.583	1.309	0.36	0.583	1.309	0.32

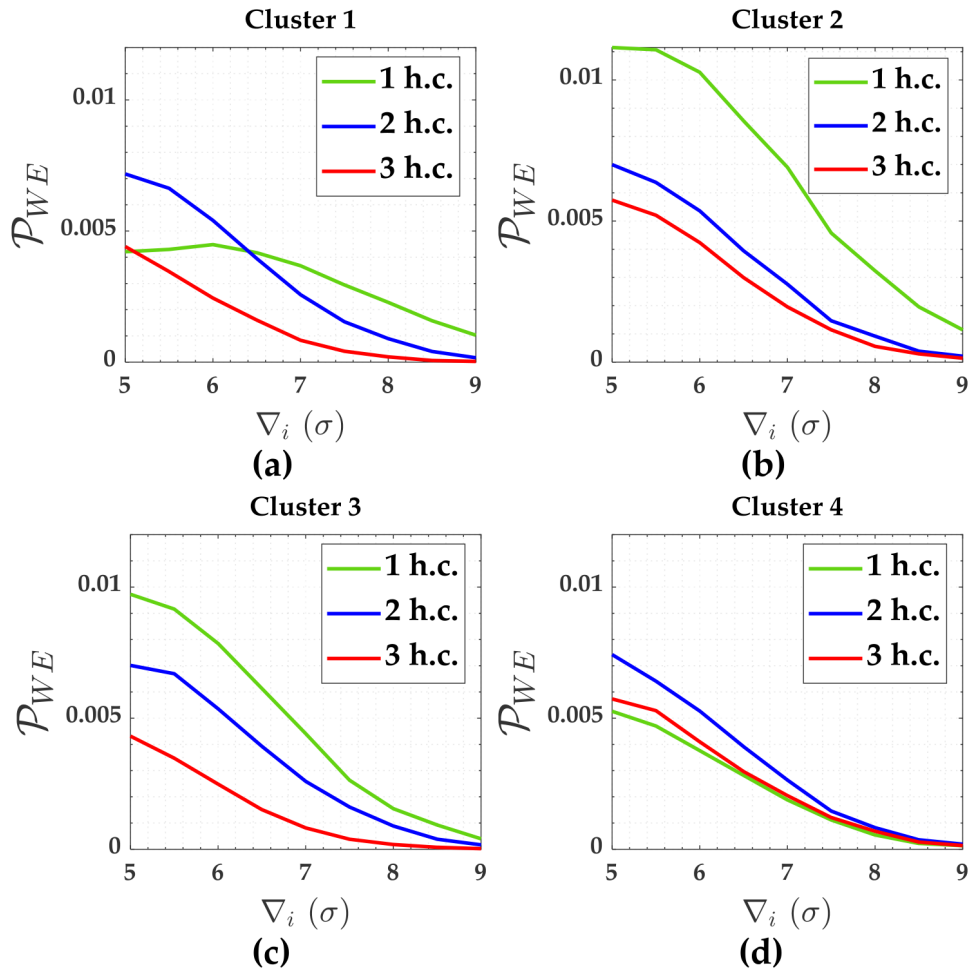


Figura 15 – Probabilidades de falsa exclusão (\mathcal{P}_{WE}) para $\hat{k} = 3.89$ no caso de uma única injunção absoluta e $\hat{k} = 3.93$ para o caso de duas e três injunções absoluta. Cluster 1 (a): y_1, y_3, y_4 e y_6 . Cluster 2 (b): y_2 e y_5 . Cluster 3 (c): y_7, y_8, y_9 e y_{10} . Cluster 4 (d): y_{11} e y_{12} .

foi ocasionada pela sobreposição estatística, como poder ser verificado pela Figura (16). Não tivemos sobreposição estatística para os clusters 2, 3 e 4.

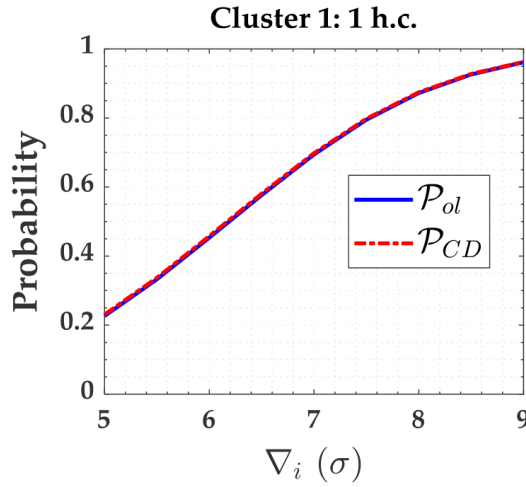


Figura 16 – Taxa de sobreposição estatística (\mathcal{P}_{ol}) e probabilidade de correta detecção (\mathcal{P}_{CD}) do Cluster 1 para o cenário de uma única injunção absoluta.

Os pontos “A” e “D” de altitude desconhecida (i.e., parâmetro no modelo) estavam conectados por apenas duas observações. Neste caso, quando o ponto “G” foi tomado como restrição rígida, os resíduos destas observações, que fazem parte do Cluster 1, apresentaram correlação igual à $\rho_{w_i, w_j} = 1.00$. Quando se têm duas observações conectadas à um ponto de altitude desconhecida, ou seja a um parâmetro, os graus de liberdade local é $gl = n - u = 2 - 1 = 1$. Se uma destas observações fossem outlier, teríamos mais um parâmetro local e, portanto, $gl = 2 - (1 + 1) = 0$. Isso significa que não teríamos graus de liberdade local suficiente para identificar o outlier, embora pudesse ser detectado. Em outras palavras, os resíduos das observações do Cluster 1 são linearmente dependentes. Por isso, a correlação entre as estatísticas de teste destas observações foram iguais à $\rho_{w_i, w_j} = 1.00$. Esta é razão da correta identificação para o Cluster 1 ter sido nula, i.e. $\mathcal{P}_{CI(Cluster\ 1)} = 0$.

As probabilidades da falsa exclusão para os casos de injunção absoluta foram menores que 1%, i.e. $\mathcal{P}_{WE} \leq 1\%$. As ocorrências de sobre-identificação (\mathcal{P}_{over+} e \mathcal{P}_{over-}) foram praticamente nulas. Como resultado, $\mathcal{P}_{CI} \approx \mathcal{P}_{CD}$. Esses resultados se deve à baixa correlação entre os resíduos ($\rho_{w_i, w_j} < 0.5$) e à alta redundância local $r_i > 0.5$. Devido à este fato, o nível de significância conjunto (α') pode ser aumentado, com o objetivo de elevar a taxa de sucesso na identificação do outlier (ROFATTO et al., 2020b).

A Figura (17) mostra uma exemplo da influência do aumento do nível de significância conjunto de ($\alpha' = 0.001$) para ($\alpha' = 0.1$) sobre os níveis de probabilidade do Cluster 1, quando sujeito à duas e três injunções absolutas.

Embora tenha aumentado as probabilidades de falsa exclusão (\mathcal{P}_{WE}) e sobre-

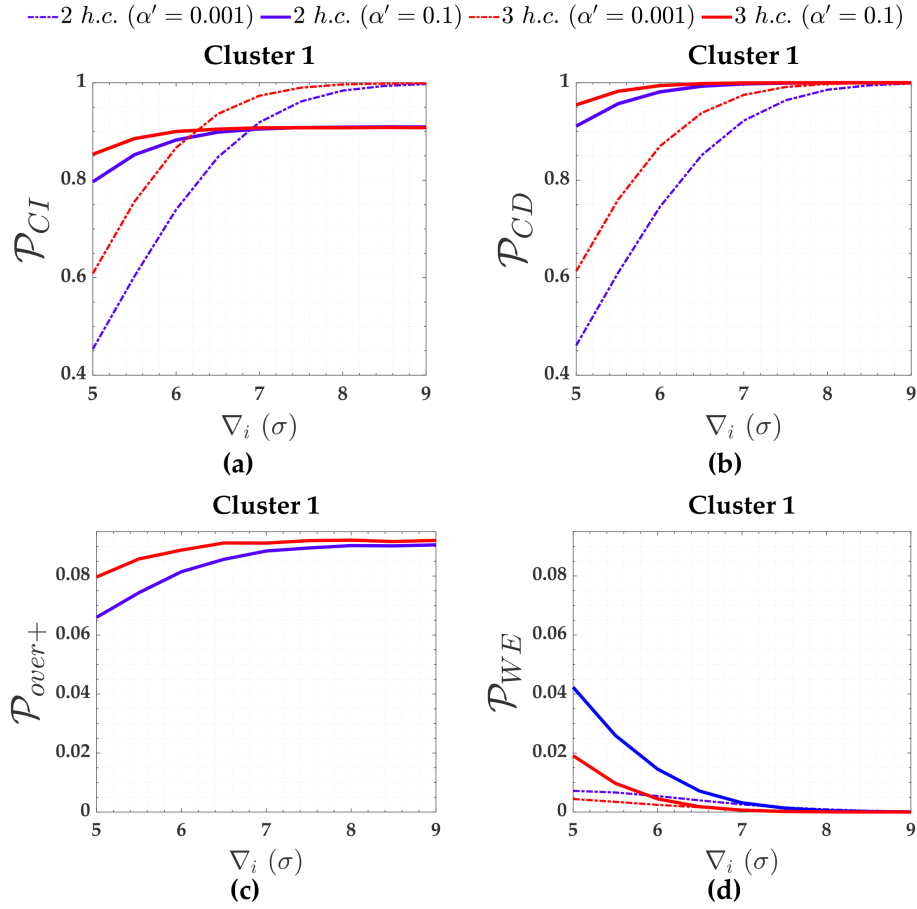


Figura 17 – Probabilidades de correta detecção (\mathcal{P}_{CD}), identificação (\mathcal{P}_{CI}), falsa exclusão (\mathcal{P}_{WE}) e sobre-identificação (\mathcal{P}_{over+}) para $\alpha' = 0.001$ e $\alpha' = 0.1$. Resultados para o Cluster 1 (y_1, y_3, y_4 e y_6) quando sujeito à duas e três inunções absolutas.

identificação (\mathcal{P}_{over+}), estas foram praticamente insignificantes quando comparadas com as melhorias das taxas de detecção (\mathcal{P}_{CD}) e identificação (\mathcal{P}_{CI}). Esta mesma análise pode ser verificada para os outros clusters.

Em termos de inunções relativas, a Tabela (3) mostra as configurações da rede para o cenário de duas inunções.

Cinco clusters foram definidos para os casos de duas inunções relativas:

- Cluster 1: y_1, y_3, y_4 e y_6 .
- Cluster 2: y_2 e y_5 .
- Cluster 3: y_7, y_8, y_9 e y_{10} .
- Cluster 4: y_{11} e y_{12} .
- Cluster 5: y_{13} (“A”) e y_{14} (“D”).

Tabela 3 – Redundância local (r_i), desvio-padrão do outlier ($\sigma_{\hat{v}_i}$) e máxima correlação absoluta entre os resíduos ($max_{\rho_{w_i, w_j}}$) para cada cenário de duas injunções relativas.

Observação	$\sigma_c = 0.1mm$			$\sigma_c = 1mm$			$\sigma_c = 10mm$		
	r_i	σ_{∇_i}	$max_{\rho_{w_i, w_j}}$	r_i	σ_{∇_i}	$max_{\rho_{w_i, w_j}}$	r_i	σ_{∇_i}	$max_{\rho_{w_i, w_j}}$
y_1	0.581	1.312	0.564	0.471	1.457	0.681	0.397	1.587	0.994
y_2	0.582	1.311	0.376	0.533	1.369	0.423	0.501	1.413	0.471
y_3	0.581	1.312	0.564	0.471	1.457	0.681	0.397	1.587	0.994
y_4	0.581	1.312	0.564	0.471	1.457	0.681	0.397	1.587	0.994
y_5	0.582	1.311	0.376	0.533	1.369	0.423	0.501	1.413	0.471
y_6	0.581	1.312	0.564	0.471	1.457	0.681	0.397	1.587	0.994
y_7	0.583	1.310	0.359	0.571	1.324	0.423	0.563	1.333	0.471
y_8	0.583	1.310	0.359	0.571	1.324	0.423	0.563	1.333	0.471
y_9	0.583	1.310	0.359	0.571	1.324	0.423	0.563	1.333	0.471
y_{10}	0.583	1.310	0.359	0.571	1.324	0.423	0.563	1.333	0.471
y_{11}	0.583	1.309	0.358	0.583	1.309	0.398	0.583	1.309	0.433
y_{12}	0.583	1.309	0.358	0.583	1.309	0.398	0.583	1.309	0.433
y_{13}	0.007	1.163	1.000	0.300	1.826	1.000	0.497	14.189	1.000
y_{14}	0.007	1.163	1.000	0.300	1.826	1.000	0.497	14.189	1.000

Para o caso de duas injunções relativas, denotadas por **s.c.** (*soft constraint*), as probabilidades de correta detecção (\mathcal{P}_{CD}) e identificação (\mathcal{P}_{CI}) das observações associadas aos clusters 1,2,3 e 4 são mostradas na Figura (18).

A Figura (19) mostra a probabilidade de correta detecção (\mathcal{P}_{CD}) para o Cluster 5, cujos elementos são as injunções relativas “A” e “D”.

Observou-se que não existe confiabilidade em termos de identificação para as injunções relativas. Esse resultado foi esperado, pois a máxima correlação entre as estatísticas das injunções foi de $max_{\rho_{w_i, w_j}} = 1.00$.

A figura (20) mostra as probabilidades de falsa exclusão (\mathcal{P}_{WE}) para os clusters das observações. Esse nível de probabilidade foi praticamente nulo para as injunções. As ocorrências de sobre-identificação (\mathcal{P}_{over+} e \mathcal{P}_{over-}) sobreposição estatística \mathcal{P}_{ol} também foram praticamente nulas neste cenário de duas injunções relativas.

Por fim, a Tabela (4) fornece as configurações para o caso de três injunções relativas. Neste caso, seis clusters foram definidos para os casos de três injunções relativas, como segue:

- Cluster 1: y_1, y_3, y_4 e y_6 .
- Cluster 2: y_2 e y_5 .
- Cluster 3: y_7, y_8, y_9 e y_{10} .
- Cluster 4: y_{11} e y_{12} .
- Cluster 5: y_{13} (“A”) e y_{14} (“D”).
- Cluster 6: y_{15} (“G”).

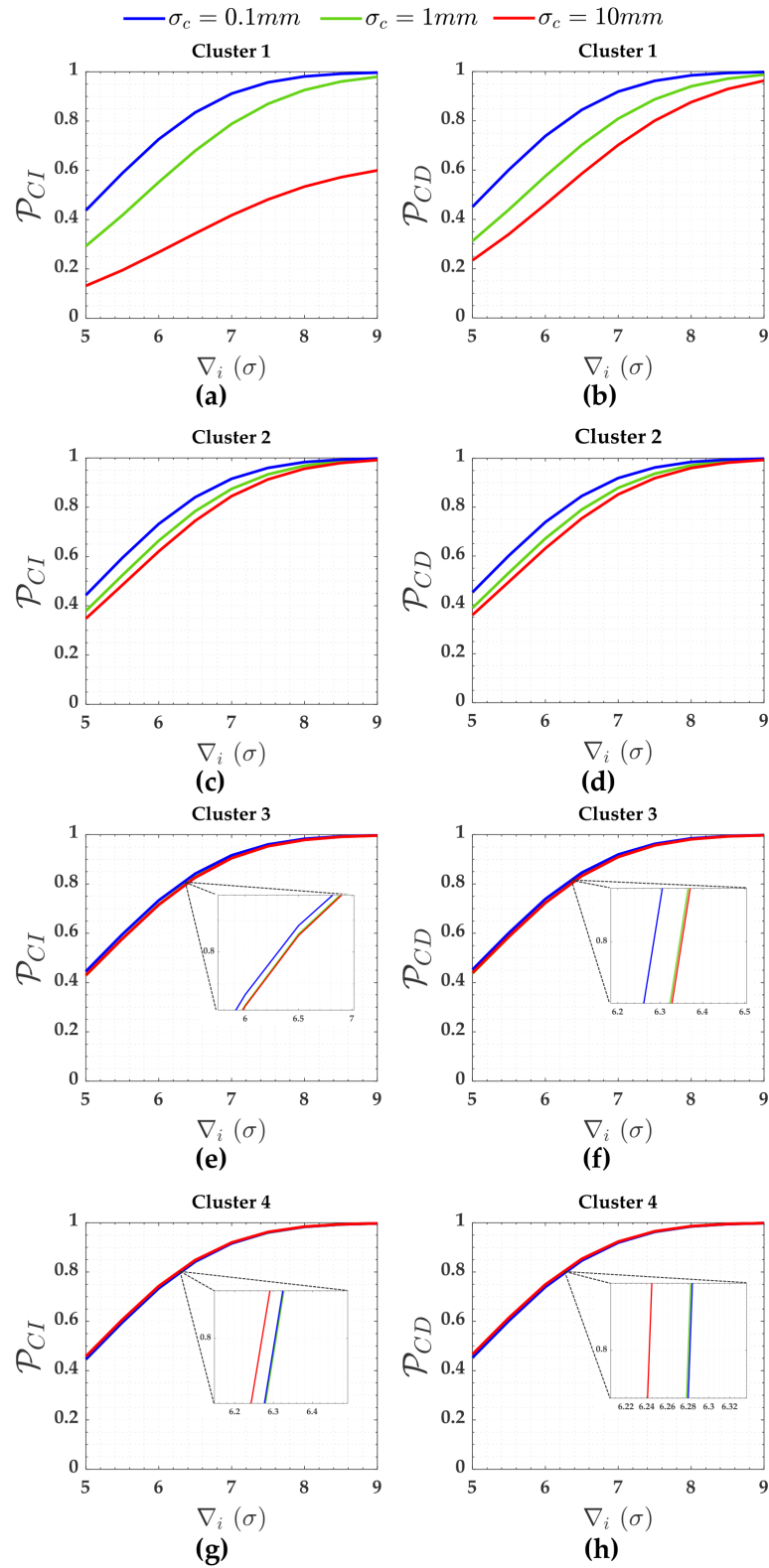


Figura 18 – Cenário para duas injunções relativas. Probabilidades de correta identificação (\mathcal{P}_{CI}) e detecção (\mathcal{P}_{CD}) para $\hat{k} = 3.95$, no caso de $\sigma_c = 0.1mm$ e $\sigma_c = 1mm$; e $\hat{k} = 3.92$ para o caso de $\sigma_c = 10mm$. Cluster 1 (a,b): y_1, y_3, y_4 e y_6 . Cluster 2 (c,d): y_2 e y_5 . Cluster 3 (e,f): y_7, y_8, y_9 e y_{10} . Cluster 4 (g,h): y_{11} e y_{12} .

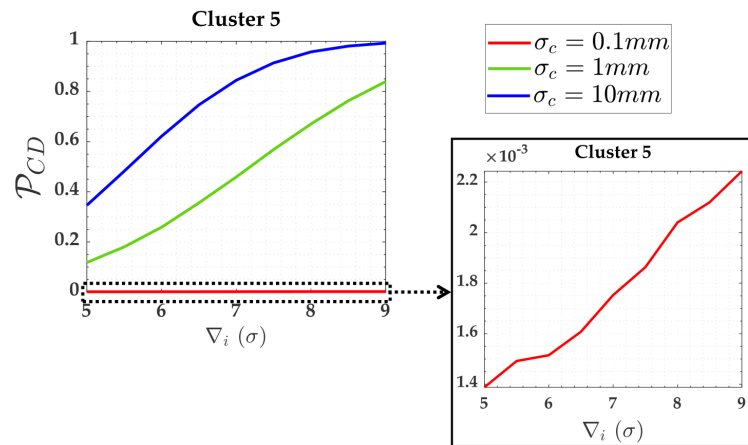


Figura 19 – Cenário para duas injeções relativas. Probabilidade de correta detecção (\mathcal{P}_{CD}) para $\hat{k} = 3.92$, no caso de $\sigma_c = 0.1mm$ e $\sigma_c = 1mm$; e $\hat{k} = 3.92$ para o caso de $\sigma_c = 10mm$. Cluster 5: y_{13} e y_{14} .

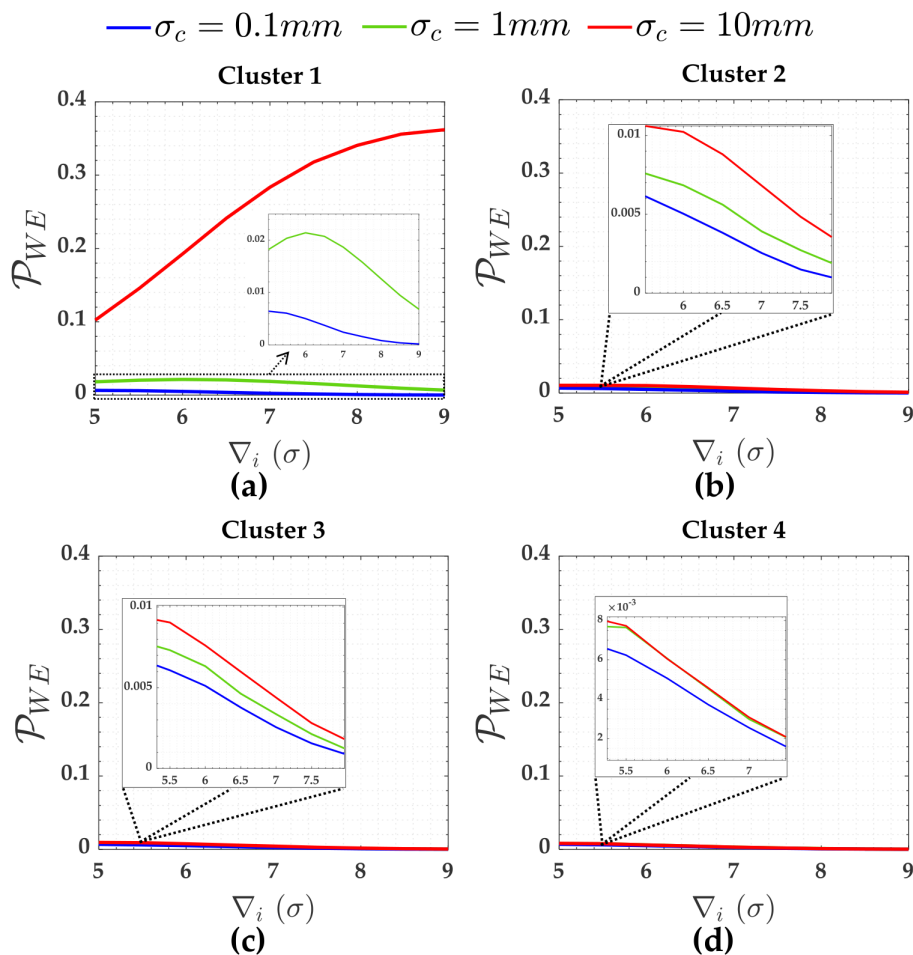


Figura 20 – Cenário para duas injeções relativas. Probabilidade de falsa exclusão (\mathcal{P}_{WE}) para $\hat{k} = 3.95$, no caso de $\sigma_c = 0.1mm$ e $\sigma_c = 1mm$; e $\hat{k} = 3.92$ para o caso de $\sigma_c = 10mm$. Cluster 1 (a): y_1, y_3, y_4 e y_6 . Cluster 2 (b): y_2 e y_5 . Cluster 3 (c): y_7, y_8, y_9 e y_{10} . Cluster 4 (d): y_{11} e y_{12} .

Tabela 4 – Redundância local (r_i), desvio-padrão do outlier (σ_{∇_i}) e máxima correlação absoluta entre os resíduos ($max_{\rho_{w_i, w_j}}$) para cada cenário de três injunções relativas.

Observação	$\sigma_c = 0.1mm$			$\sigma_c = 1mm$			$\sigma_c = 10mm$		
	r_i	σ_{∇_i}	$max_{\rho_{w_i, w_j}}$	r_i	σ_{∇_i}	$max_{\rho_{w_i, w_j}}$	r_i	σ_{∇_i}	$max_{\rho_{w_i, w_j}}$
y_1	0.702	1.194	0.660	0.502	1.411	0.577	0.398	1.586	0.992
y_2	0.582	1.311	0.326	0.533	1.369	0.412	0.501	1.413	0.470
y_3	0.702	1.194	0.660	0.502	1.411	0.577	0.398	1.586	0.992
y_4	0.702	1.194	0.660	0.502	1.411	0.577	0.398	1.586	0.992
y_5	0.582	1.311	0.326	0.533	1.369	0.412	0.501	1.413	0.470
y_6	0.702	1.194	0.660	0.502	1.411	0.577	0.398	1.586	0.992
y_7	0.704	1.192	0.415	0.602	1.289	0.412	0.563	1.333	0.470
y_8	0.704	1.192	0.415	0.602	1.289	0.412	0.563	1.333	0.470
y_9	0.704	1.192	0.415	0.602	1.289	0.412	0.563	1.333	0.470
y_{10}	0.704	1.192	0.415	0.602	1.289	0.412	0.563	1.333	0.470
y_{11}	0.583	1.309	0.326	0.583	1.309	0.385	0.583	1.309	0.433
y_{12}	0.583	1.309	0.326	0.583	1.309	0.385	0.583	1.309	0.433
y_{13}	0.012	0.904	0.660	0.425	1.534	0.542	0.663	12.283	0.501
y_{14}	0.012	0.904	0.660	0.425	1.534	0.542	0.663	12.283	0.501
y_{15}	0.019	0.718	0.63	0.5	1.414	0.542	0.665	12.268	0.501

As figuras (21) e (22) mostra as probabilidades de correta identificação (\mathcal{P}_{CI}) e detecção (\mathcal{P}_{CD}) para os clusters das observações (1, 2, 3, 4) e para os clusters das injunções (Cluster 5: “A” e “D”; Cluster 6: “G”), respectivamente. A figura (23) mostra as probabilidades de falsa exclusão (\mathcal{P}_{WE}) para os clusters das observações.

As ocorrências de sobre-identificação (\mathcal{P}_{over+} e \mathcal{P}_{over-}) sobreposição estatística \mathcal{P}_{ol} , bem como a taxa de falsa exclusão (\mathcal{P}_{WE}) foram praticamente nulas neste cenário de três injunções relativas.

Com base nos níveis de probabilidade estimados, os indicadores de MDB e MIB foram obtidos para cada cenário de injunção. Os valores de MDB e MIB foram calculados para $\alpha' = 0.001$ e $\tilde{\mathcal{P}}_{CD} = \tilde{\mathcal{P}}_{CI} = 80\%$, de acordo com as expressões (3.18) e (3.19), respectivamente. As Tabelas (5), (6) e (7) mostram os indicadores de sensibilidade do DSI para os cenários de injunção absoluta, relativa com duas injunções e relativa com três injunções, respectivamente.

Os resultados revelaram os seguintes aspectos:

- Quanto maior o número de injunções absolutas, maior é a probabilidade da correta identificação \mathcal{P}_{CI} .
- O cenário de adotar uma única restrição flexível não foi considerado, pois recairia sobre o problema de não ter graus de liberdade suficiente nem mesmo para detectar outlier. Portanto, o mínimo necessário são duas injunções relativas.
- Os resultados mostraram que quanto maior é a *relaxação* da restrição ($\sigma_c \gg 0$), maior é o efeito da correlação entre os resíduos sobre a taxa de correta identificação.

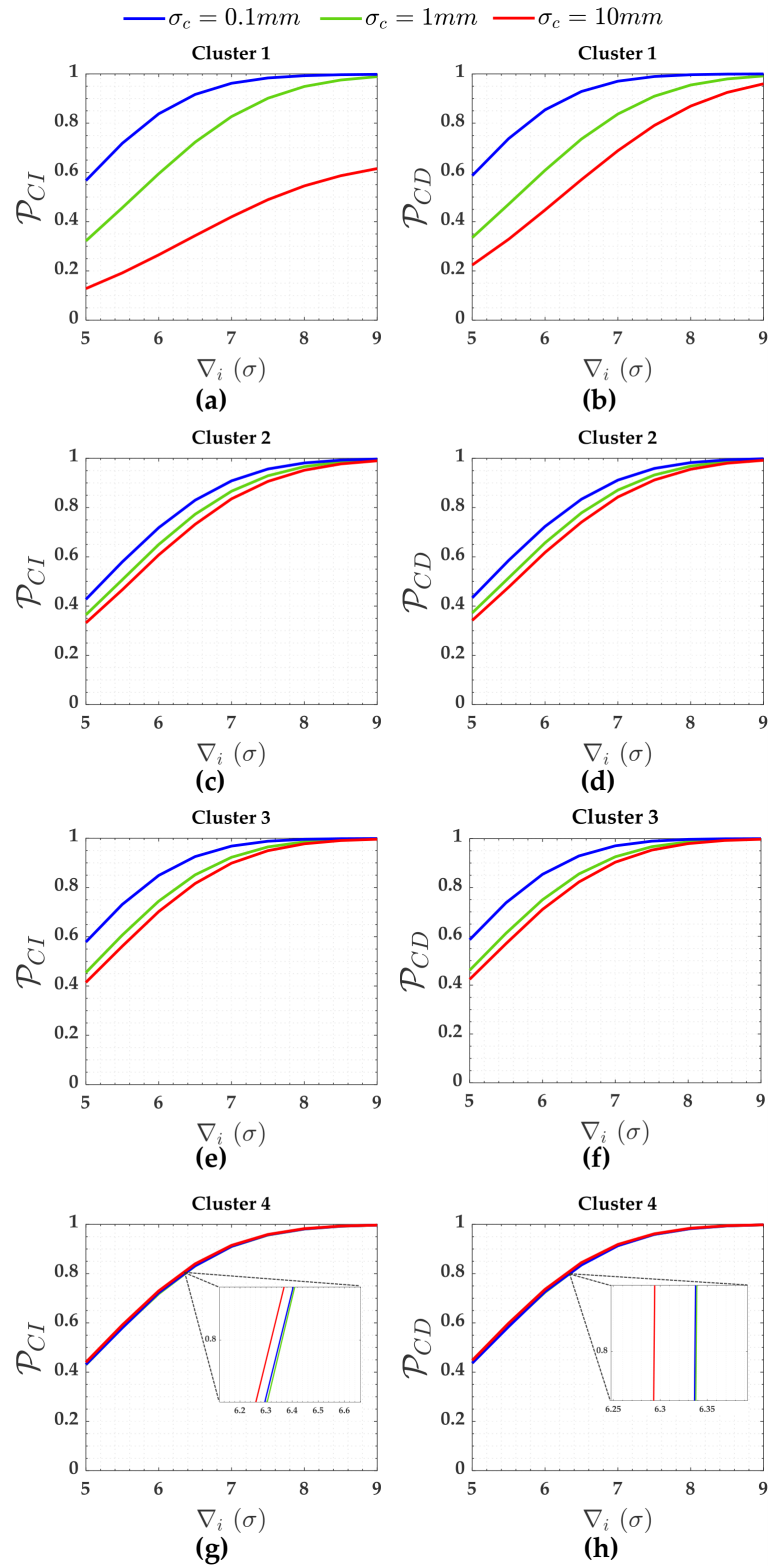


Figura 21 – Cenário para três injunções relativas. Probabilidades de correta identificação (\mathcal{P}_{CI}) e detecção (\mathcal{P}_{CD}) para $\hat{k} = 3.99$, no caso de $\sigma_c = 0.1mm$ e $\sigma_c = 1mm$; e $\hat{k} = 3.96$ para o caso de $\sigma_c = 10mm$. Cluster 1 (a,b): y_1, y_3, y_4 e y_6 . Cluster 2 (c,d): y_2 e y_5 . Cluster 3 (e,f): y_7, y_8, y_9 e y_{10} . Cluster 4 (g,h): y_{11} e y_{12} .

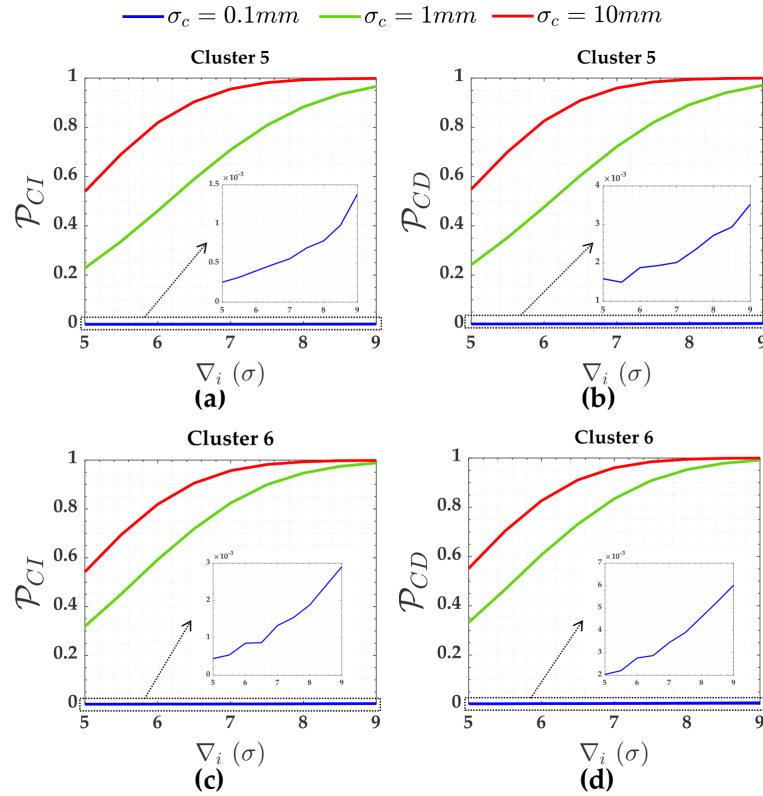


Figura 22 – Cenário para três injeções relativas. Probabilidades de correta identificação (\mathcal{P}_{CI}) e detecção (\mathcal{P}_{CD}) para $\hat{k} = 3.99$, no caso de $\sigma_c = 0.1mm$ e $\sigma_c = 1mm$; e $\hat{k} = 3.96$ para o caso de $\sigma_c = 10mm$. Cluster 5 (a,b): y_{13} e y_{14} . Cluster 6 (c,d): y_{15} .

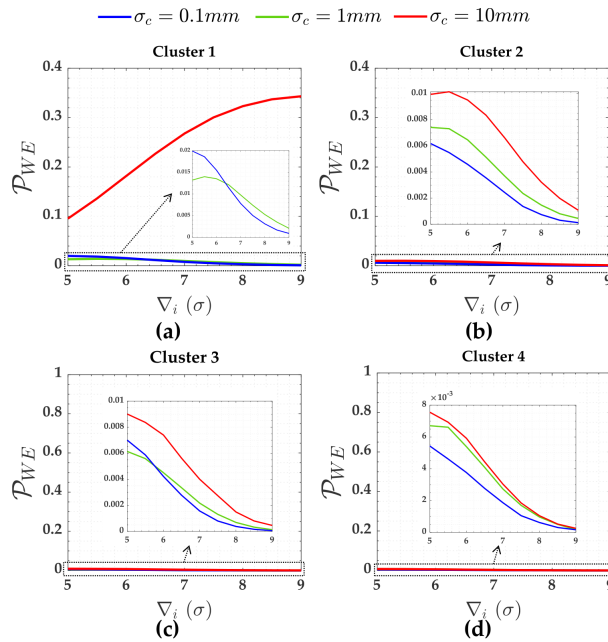


Figura 23 – Cenário para três injeções relativas. Probabilidade de falsa exclusão (\mathcal{P}_{WE}) para $\hat{k} = 3.99$, no caso de $\sigma_c = 0.1mm$ e $\sigma_c = 1mm$; e $\hat{k} = 3.96$ para o caso de $\sigma_c = 10mm$. Cluster 1 (a,b): y_1, y_3, y_4 e y_6 . Cluster 2 (c,d): y_2 e y_5 . Cluster 3 (e,f): y_7, y_8, y_9 e y_{10} . Cluster 4 (g,h): y_{11} e y_{12} .

Conseqüentemente, maior é o indicador de sensibilidade para fins de identificação de outlier (MIB) e, portanto, mais difícil se torna identificar um outlier.

- Em geral, a probabilidade da correta identificação \mathcal{P}_{CI} para o caso de injunções absolutas é maior que a das injunções relativas.

Tabela 5 – MDB e MIB para o caso de injunções absolutas com $\alpha' = 0.001$ e $\tilde{\mathcal{P}}_{CD} = \tilde{\mathcal{P}}_{CI} = 0.8$.

Cluster	1 inj. absoluta		2 inj. absolutas		3 inj. absolutas	
	MDB (σ)	MIB (σ)	MDB (σ)	MIB (σ)	MDB (σ)	MIB (σ)
1	7.5	-	6.3	6.3	5.7	5.7
2	6.7	6.8	6.3	6.4	6.3	6.4
3	6.4	6.4	6.3	6.3	5.8	5.8
4	6.4	6.4	6.4	6.4	6.4	6.4

Tabela 6 – MDB and MIB para o caso de duas injunções relativas com $\alpha' = 0.001$ e $\tilde{\mathcal{P}}_{CD} = \tilde{\mathcal{P}}_{CI} = 0.8$.

Cluster	$\sigma_c = 10mm$		$\sigma_c = 1mm$		$\sigma_c = 0.1mm$	
	MDB (σ)	MIB (σ)	MDB (σ)	MIB (σ)	MDB (σ)	MIB (σ)
1	7.5	25	7	7.1	6.3	6.3
2	6.8	6.8	6.6	6.6	6.3	6.3
3	6.4	6.4	6.4	6.4	6.3	6.3
4	6.3	6.3	6.3	6.3	6.3	6.3
5	6.8	-	8.8	-	57	-

Tabela 7 – MDB and MIB para o caso de três injunções relativas com $\alpha' = 0.001$ e $\tilde{\mathcal{P}}_{CD} = \tilde{\mathcal{P}}_{CI} = 0.8$.

Cluster	$\sigma_c = 10mm$		$\sigma_c = 1mm$		$\sigma_c = 0.1mm$	
	MDB (σ)	MIB (σ)	MDB (σ)	MIB (σ)	MDB (σ)	MIB (σ)
1	7.5	22	6.8	6.9	5.8	5.9
2	6.8	6.9	6.6	6.7	6.4	6.4
3	6.4	6.4	6.3	6.3	5.8	5.8
4	6.3	6.3	6.3	6.3	6.3	6.3
5	5.9	6.0	7.4	7.5	43.5	45
6	5.9	5.9	6.9	6.9	34.6	35.5

On the Effects of Hard and Soft Equality Constraints in the Iterative Outlier Elimination Procedure

Vinicius Francisco Rofatto^{1,2*}, Marcelo Tomio Matsuoka^{1,2,3}, Ivandro Klein^{4,5}, Maurício Roberto Veronez⁶, Luiz Gonzaga da Silveira Junior⁶

1 Graduate Program in Remote Sensing, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

2 Institute of Geography, Federal University of Uberlândia, Monte Carmelo, MG, Brazil

3 Graduate Program in Agriculture and Geospatial Information, Federal University of Uberlândia, Monte Carmelo, MG, Brazil

4 Department of Civil Construction, Federal Institute of Santa Catarina, Florianópolis, SC, Brazil

5 Graduate Program in Geodetic Sciences, Federal University of Paraná, Curitiba, PR, Brazil

6 Graduate Program in Applied Computing, Unisinos University, São Leopoldo, RS, Brazil

* vfrofatto@gmail.com

Abstract

The reliability analysis allows to estimate the system's probability of detecting and identifying outlier. Failure to identify an outlier can jeopardise the reliability level of a system. Due to its importance, outliers must be appropriately treated to ensure the normal operation of a system. The system models are usually developed from certain constraints. Constraints play a central role in model precision and validity. In this work, we present a detailed optical investigation of the effects of the hard and soft constraints on the reliability of a measurement system model. Hard constraints represent a case in which there exist known functional relations between the unknown model parameters, whereas the soft constraints are employed for the case where such functional relations can slightly be violated depending on their uncertainty. The results highlighted that the success rate of identifying an outlier for the case of hard constraints is larger than soft constraints. This suggested that hard constraints be used in the stage of pre-processing data for the purpose of identifying and removing possible outlying measurements. After identifying and removing possible outliers, one should set up the soft constraints to propagate their uncertainties to the model parameters during the data processing.

Introduction

It is very common to build models (i.e. the equation systems) based on some initial knowledge about a given problem. In other words, models are often set up in a way that the model parameters need to fulfil certain constraints. Such constraints are a prior knowledge embedded into a model to avoid a trivial solution; to guarantee the stability of estimates; to improve the precision and accuracy of the results by reducing the number of unknown parameters or accordingly by increasing the redundancy of the system; and to mitigate (or even estimate) a possible systematic effect [1, 2].

The models are usually formulated with minimal constraint or extra (redundant) constraints. In that case, we are referring to the so-called *equality constraints* which are usually incorporated into a system of equations in order to try to have a model well posed [3]. For the most part, minimal constraints are introduced to solve the problem of rank deficiency in linear (or linearised) systems. The rank deficiency is often caused by the lack (or insufficient) information about a problem. In the field of geodesy, for example, minimal constraints are external information whose primary role is to specify the coordinate system to which the network station positions will be estimated by least-squares method (LS). This problem is known as *datum definition* (or also zero-order design or datum choice problem) [4–9]. Several works have been investigated in the sense of minimum-constrained adjustment and the datum choice problem in the geodetic literature, focusing on topics like free-adjustment and the role of inner constraints [10–13].

If the number of constraints exceeds the minimum needed to solve the rank deficiency of the equation systems, we say that we have redundant (or extra) constraints. Extra constraints have also been used to check the stability of points in geodetic deformation analysis [14–16], to test the compatibility of constraints with the observations and the rest of the constraints [17–19].

So far we have only distinguished the constraints in terms of numerical quantity. The model can also be subject to a *hard* and *soft* (or *weighted*) constraints. Hard constraints can often represent a case in which there exist known functional relations between the unknown parameters. Soft constraints (or *looser* constraints) are, however, for which such functional relations can slightly be violated depending on their uncertainty [2, 19]. Soft constraints may also be referred to as a pseudo-observation model [20].

The well-known least-squares (LS) has been widely used as a standard method of model parameters estimation in geodetic applications and many others branches of modern science [21–41]. This is due to the flexibility of the LS, since no concepts from probability theory are used in formulating the least-squares minimisation problem.

LS has the property of being a linear unbiased estimator (LUE), and some special cases it coincides with the best linear unbiased estimator (BLUE). The estimator which has the smallest variance of all LUEs is called the best linear unbiased estimator (BLUE). If we have the full knowledge of the probability density function (PDF) of the measurements, the method of maximum likelihood estimation (MLE) can also be applied. In case of normally distributed measurements (Gauss-Markov model), the MLE estimators are identical to the BLUE ones, and therefore the LS and MLE principles provide identical results [24, 42]. However, the presence of undesirable outliers in dataset makes LS no longer *unbiased* and it does not coincides with MLE [43].

Here, we assume that outlier is an observation that has deviated from its most probable value to the point of jeopardising the mathematical model (functional and stochastic) to which it should belong. Due to its importance, outliers must be appropriately treated to ensure the quality of data analysis [44–50].

In this paper, we employ the iterative data snooping (*IDS*), which is hypothesis test-based outlier. Important to mention that *IDS* is not restrict to the field of geodetic statistics, but it is a generally applicable method [51, 52]. *IDS* is an iterative outlier elimination procedure, which combines estimation, testing and a corrective action [44, 53]. Parameter estimation is often conducted in the sense of the LS. Then, hypothesis testing is performed with the aim to identify any outlier that may be present in dataset. After identification, the suspected outlier is then excluded from the dataset as a corrective action (i.e. adaptation), and the LS is restarted without the rejected measurement. If the model redundancy permits, this procedure is repeated until no more (possible) outliers can be identified (see e.g. [23], pp. 135). Although here we restrict ourselves to the case of one outlier at a time, *IDS* can also be applied for the

case of multiples (simultaneous) outliers [54]. For more details about multiples (simultaneous) outlier refers to [55–57]. Because *IDS* is based on statistical hypothesis testing, there are chances of correct and false decisions. Recently, Rofatto et al. [44] have provided an algorithm based on Monte Carlo to determine the probability levels associated with *IDS*. In that case, they described six classes of decisions for *IDS*, namely probability of correct identification (\mathcal{P}_{CI}), probability of missed detection (\mathcal{P}_{MD}), probability of wrong exclusion (\mathcal{P}_{WE}), probability of over-identification positive (\mathcal{P}_{over+}), probability of over-identification negative (\mathcal{P}_{over-}) and probability of statistical overlap (\mathcal{P}_{ol}), as follows:

- \mathcal{P}_{CI} : Probability of identifying and removing correctly an outlying measurement;
- \mathcal{P}_{MD} : Probability of not detecting the outlier (i.e. Type II decision error for *IDS*);
- \mathcal{P}_{WE} : Probability of identifying and removing a non-outlying measurement while the ‘true’ outlier remains in the dataset (i.e. Type III decision error [58] for *IDS*);
- \mathcal{P}_{over+} : Probability of identifying and removing correctly the outlying measurement and others;
- \mathcal{P}_{over-} : Probability of identifying and removing more than one non-outlying measurement, whereas the ‘true outlier’ remains in the dataset;
- \mathcal{P}_{ol} : occurs in cases where one alternative hypothesis has the same distribution as the another one. These hypotheses cannot be distinguished, because their test statistics are numerically the same, violating the *IDS* rule of one outlier at a time. In that case, they are nonseparable and an outlier cannot be identified. In other words, it corresponds to the probability of flagging simultaneously two (or more) measurements as outliers.

Based on the probability of correct detection (\mathcal{P}_{CD}) and probability of correct identification (\mathcal{P}_{CI}), the minimal biases, MDB (Minimal Detectable Bias) and MIB (Minimal Identifiable Bias), can be computed as sensitivity indicators for outlier detection and identification, respectively. “Outlier Detection” only informs whether or not there might have been at least one outlier. However, the detection does not tell us which measurement is an outlier. The localization of the outlier is a problem of “outlier identification”, i.e. “Outlier Identification” implies the execution of a search among the measurements for the most likely outlier [44]. Therefore, the smallest value of an outlier that can be detected given a certain \mathcal{P}_{CD} defines the MDB. On the other hand, the smallest value of an outlier that can be identified given a certain \mathcal{P}_{CI} defines the MIB.

In this paper, we investigate the effects of models subject to constraints (minimum, redundant, hard and soft) on the probability levels associated with *IDS*. It is important to emphasised that if a standard-deviation of a constraint (or a set of a constraint) is changed from zero to a nonzero value, it is called “relaxation” of the constraint [20].

Here, we also evaluate the effect of relaxing constraints on the MIB and MDB. This kind of assessment is a kind of sensitivity analysis. We also highlight that the task of clustering a set of geodetic measurements is applied for the first time in this study. We show that the clusters can be defined according to two deterministic parameters: local redundancy and correlation between the outlier test statistics.

Moreover, critical values optimized by Monte Carlo method were used here [44, 51] in order to compute the decision classes associated with *IDS*, i.e. \mathcal{P}_{CI} , \mathcal{P}_{MD} , \mathcal{P}_{WE} , \mathcal{P}_{over+} , \mathcal{P}_{over-} and \mathcal{P}_{ol} .

Material and methods

Here, we use the procedure provided by Rofatto et al. [44] to compute the probability levels associated with *IDS*, as well as to estimate the minimal biases — Minimal Detectable Bias (MDB) and Minimal Identifiable Bias (MIB). The procedure is summarised in Fig 1.

Fig 1. Flowchart of the algorithm. Flowchart of the algorithm to compute the probability levels of Iterative Data Snooping (IDS) for each measurement in the presence of an outlier [44].

Here, the probability levels associated with *IDS* are computed for each observation individually and for each outlier magnitude. However, they are grouped into clusters based on number of local redundancy (r_i) and maximum absolute correlation between the outlier test statistics (ρ_{w_i, w_j}). Furthermore, we take care to control the *family-wise error rate*. See Supporting Information for more details S1 Appendix.

Problem Description

To analyse the constraints effects on the *IDS*, an example is taken from a geodetic levelling network with 12 height differences between the points. The equipment used to measure the level difference can be an electronic digital level. In that case, the levelling measurement system comprises of a special bar-coded staff (also called barcode rod) and a digital level (instrument). A digital level is basically a telescope that enables a horizontal line of sight. Digital levels consist of additional electronic image processing components to automatically read and analyse digital (bar coded) levelling staffs, where the graduation is replaced by a manufacturer dependent code pattern. Generally, the result is automatically stored in the data collector of the digital level. An example of a "digital level – bar-code staff" system is displayed in the Fig 2. For more details about digital level see e.g. [59–61].

Fig 2. digital level – bar-code staff system. Example of a *digital level – bar-code staff* system [44].

The standard-deviation of the uncorrelated measurements were the same and taken equal to $\sigma = 1mm$. The points are indicated as A to G. The eight network configuration are displayed in Fig 3(a, b, c, d, and e) and their details are given as follows:

1. Figure 3(a): Network with 1 hard constraint (i.e. network minimally constrained). Since the dimension of the network is 1D, the minimum information necessary to estimate the unknown heights is one. The height of G was fixed as control point (hard constraint), and 6 unknown heights (A,B,C,D,E,F) were minimally constrained. Therefore, the redundancy of the system (or overall degrees of freedom) was $r = n - rank(A) = n - u = 12 - 6 = 6$.
2. Figure 3(b): Network with one extra hard constraint (i.e. two hard constraints). The heights A and D were taken as hard constraints (i.e. heights A and D were fixed). The redundancy of the system in that case was $r = 12 - 5 = 7$ with 5 unknown heights (B,C,E,F,G) over-constrained.
3. Figure 3(c): Network with two extra hard constraints (i.e. three hard constraints). The heights A, D and G were taken as hard constraints. In that case, the redundancy of the system was $r = 12 - 4 = 8$.

4. Figure 3(d): Network with two soft constraints (A and D). In that case, a standard-deviation larger than zero was assigned to both constraints (i.e. $\sigma_c > 0$). In other words, A and D were processed as being both observations and unknown parameters, i.e. A and D were pseudo-observations. Here, the both constraints were simultaneously *relaxed* by considering their uncertainties 10 times worse than measurements (i.e. $\sigma_c = 10 \times \sigma = 10mm$); 10 times better than measurements (i.e. $\sigma_c = 0.1mm$); and their uncertainties equal to the measurements ($\sigma_c = 1mm$). In that case, the redundancy of the system was $r = 14 - 7 = 7$.
5. Figure 3(e): Network processed with A, D and G as pseudo-observations. Those three constraints were simultaneously *relaxed* by considering their standard-deviations equal to $\sigma_c = 10mm$ (10 times worse than measurements); $\sigma_c = 0.1mm$ (10 times better than measurements); and $\sigma_c = 1mm$ (the same as the measurements). In that case, the redundancy of the system was $r = 15 - 7 = 8$.

Fig 3. Different constraint scenarios. Levelling Geodetic Network subject to different constraint scenarios.

The following system of equations for that problem is given by:

$$\begin{aligned}
y_1 + e_1 &= h_B - h_A \\
y_2 + e_2 &= h_C - h_B \\
y_3 + e_3 &= h_D - h_C \\
&\vdots \\
y_7 + e_7 &= h_B - h_G \\
y_8 + e_8 &= h_C - h_G \\
&\vdots \\
y_{11} + e_{11} &= h_B - h_F \\
y_{12} + e_{12} &= h_C - h_E
\end{aligned} \tag{1}$$

The design matrix (\mathbf{A}) for the system of equations in 1 is given by:

$$\mathbf{A} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 \end{bmatrix} \tag{2}$$

Note that the rank defect of the matrix \mathbf{A} is $u - \text{rank}(\mathbf{A}) = 7 - 6 = 1$. In that case, there is needed to add at least one constraint in order to avoid rank deficiency of the matrix \mathbf{A} . This is guaranteed when we take one height as known. For example, from the network in Figure 3(a), we have added the height G as known (i.e. as a hard constraint). In that case, the constraint equation should be added into the system in 1, i.e.:

$$y_{13} = h_G \text{ with } \sigma_{y_{13}} = 0, \quad (3)$$

noticing that because the standard deviation is zero, the observation is non-stochastic (hard constraint) and the residual $e_{y_{13}} = 0$. This can generate problems in the inversion of the covariance matrix of the observations \mathbf{Q}_e for the calculation of the weight matrix \mathbf{W} , because the weight for that constraint would be $\frac{1}{0} = \infty$. In order to avoid that problem, we have eliminated the rank deficiency of matrix \mathbf{A} by removing the seventh column of matrix \mathbf{A} in 2 associated with the height G. Now, we have $u - \text{rank}(A) = 6 - 6 = 0$. The constraint defines the geodetic datum, i.e. the S-system [62]. Another approach to solving the system of equations in 1 could be based on generalised (pseudo) inverses [63].

The location of the constraints can be chosen in some circumstances, for example during the design stage of a geodetic network. For the special case of having a minimally constrained system, the location of the constraint will not influence the w -test statistics and the sensitivity indicators (MIB and MDB) [9]. However, more constraints than the minimum necessary to have a solution (i.e. extra constraints or redundant constraints) can change the least-squares residuals and hence w -test statistics and the minimal biases.

From the network with one extra constraint (2 constraints) in Figure 3(b), for example, the both first (height A) and fourth column (height D) of matrix \mathbf{A} in 2 were eliminated in the case of having the two heights as hard constraints. For the case where these two heights (A and D) were taken as soft constraints, however, two observation equations were added to Eq 1, i.e.:

$$\begin{aligned} y_{13} + e_{13} &= h_A, \quad \sigma_{y_{13}} > 0 \\ y_{14} + e_{14} &= h_D, \quad \sigma_{y_{14}} > 0 \end{aligned} \quad (4)$$

In the case of soft constraints in Eq 4, two lines were added in matrix \mathbf{A} . In other words, A and D were taken as pseudo-observations. In that case, the rank deficiency was also null (i.e. $u - \text{rank}(A) = 7 - 7 = 0$), the redundancy of the system was $r = n - \text{rank}(A) = n - u = 7$ and the matrix \mathbf{A} was given as follows:

$$\mathbf{A} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (5)$$

For this example of two soft constraints, and by considering the both soft constraints with standard-deviation $\sigma_c = 10mm$, the symmetric and positive semi-definite covariance matrix of the observations (\mathbf{Q}_e) was given as follows:

$$\mathbf{Q}_e = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 100 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 100 \end{bmatrix} \quad (6)$$

The last two rows and columns of the matrix \mathbf{Q}_e in 6 refer to the variances ($\sigma_c^2 = (10mm)^2 = 100mm^2$) of the heights constraints A and D, respectively. Similarly, matrices \mathbf{A} and \mathbf{Q}_e were constructed for the other cases studied here.

Although the measurements are able to identify an outlier for the case of having only one single soft constraint, the pseudo-observation (constraint) is not. In that case, the defect configuration is associated with the additional parameter in the constraint (i.e. the presence of an outlier in the constraint). In other words, an additional parameter on the soft constraint will not be estimable. For example, if the height point G was taken as a soft constraint, the presence of an outlier in pseudo-observation G would lead to rank deficiency of matrix \mathbf{A} , i.e. $u - rank(A) = 8 - 7 = 1$. Therefore, the case of having only one single soft constraint was not considered here.

Result of the Hard Constraint Effects on the Iterative Outlier Elimination Procedure

The scenarios in Figure 3(a) (network minimally constrained), Figure 3(b) (two hard constraints) and Figure 3(c) (three hard constraints) were considered here for the analysis. Table 1 gives the local redundancy (r_i), the standard-deviation of the LS-estimated outlier σ_{∇_i} and the maximum absolute correlation ($max_{\rho_{w_i, w_j}}$) for each scenario of hard constraint set out of in this study, i.e. Figure 3(a,b and c).

Table 1. Local redundancy (r_i), standard-deviation of the LS-estimated outlier σ_{∇_i} and the maximum absolute correlation ($max_{\rho_{w_i, w_j}}$) for each scenario of hard constraint.

Measurement	1 hard constraint			2 hard constraints			3 hard constraints		
	r_i	σ_{∇_i}	$max_{\rho_{w_i, w_j}}$	r_i	σ_{∇_i}	$max_{\rho_{w_i, w_j}}$	r_i	σ_{∇_i}	$max_{\rho_{w_i, w_j}}$
y_1	0.396	1.589	1.00	0.583	1.309	0.36	0.708	1.188	0.41
y_2	0.500	1.414	0.47	0.583	1.309	0.36	0.583	1.309	0.32
y_3	0.396	1.589	1.00	0.583	1.309	0.36	0.708	1.188	0.41
y_4	0.396	1.589	1.00	0.583	1.309	0.36	0.708	1.188	0.41
y_5	0.500	1.414	0.47	0.583	1.309	0.36	0.583	1.309	0.32
y_6	0.396	1.589	1.00	0.583	1.309	0.36	0.708	1.188	0.41
y_7	0.563	1.333	0.47	0.583	1.309	0.36	0.708	1.188	0.41
y_8	0.563	1.333	0.47	0.583	1.309	0.36	0.708	1.188	0.41
y_9	0.563	1.333	0.47	0.583	1.309	0.36	0.708	1.188	0.41
y_{10}	0.563	1.333	0.47	0.583	1.309	0.36	0.708	1.188	0.41
y_{11}	0.583	1.309	0.43	0.583	1.309	0.36	0.583	1.309	0.32
y_{12}	0.583	1.309	0.43	0.583	1.309	0.36	0.583	1.309	0.32

Next, the twelve levelling measurements were clustered into four clusters. The four clusters were defined as follows:

- Cluster 1: y_1, y_3, y_4 and y_6 .
- Cluster 2: y_2 and y_5 .

- Cluster 3: y_7, y_8, y_9 and y_{10} .
- Cluster 4: y_{11} and y_{12} .

The probability levels associated with *IDS* were averaged for each of these clusters. The critical values were $\hat{k} = 3.89$, $\hat{k} = 3.93$ and $\hat{k} = 3.93$ for 1 hard constraint, 2 hard constraints and 3 hard constraints, respectively. These critical values were found for $\alpha' = 0.001$. The probability of correct identification (\mathcal{P}_{CI}) and correct detection ($\mathcal{P}_{CD} = 1 - \mathcal{P}_{MD}$) are displayed in Figure 4 for each number of hard constraint (denoted by **h.c.**).

Fig 4. Probability of correct identification (\mathcal{P}_{CI}) and Probability of correct detection (\mathcal{P}_{CD}) for the case of hard constraints and for $\alpha' = 0.001$. Cluster 1(a,b), Cluster 2(c,d), Cluster 3(e,f) and Cluster 4(g,h).

The outlier magnitude were defined from $|5\sigma|$ to $|9\sigma|$. The outlier of $|5\sigma|$ was chosen because it is approximately the lowest $MDB_{0(i)}$ of the network when a single hypothesis testing is in play (See supplementary material for more details S1 Appendix). That $MDB_{0(i)}$ of $|5\sigma|$ was computed for significance level of $\alpha' = 0.001$ and a power of the test $\gamma_0 = 0.8$. This strategy reduces the search space for a MIB (Minimal Identifiable Bias), because we will always have the following inequality $MIB \geq MDB_{0(i)}$ [52, 64]. Remember that the *IDS* procedure is an example of multiple hypothesis testing. The success rate for outlier detection and outlier identification were taken as being $\tilde{\mathcal{P}}_{CD} = \tilde{\mathcal{P}}_{CI} = 0.8$, respectively. Table 2 provides the values of MDB and MIB for that case of hard constraints.

Table 2. MDB and MIB for the case of hard constraints based on $\alpha' = 0.001$ and $\tilde{\mathcal{P}}_{CD} = \tilde{\mathcal{P}}_{CI} = 0.8$.

Cluster	1 hard constraint		2 hard constraints		3 hard constraints	
	MDB (σ)	MIB (σ)	MDB (σ)	MIB (σ)	MDB (σ)	MIB (σ)
1	7.5	-	6.3	6.3	5.7	5.7
2	6.7	6.8	6.3	6.4	6.3	6.4
3	6.4	6.4	6.3	6.3	5.8	5.8
4	6.4	6.4	6.4	6.4	6.4	6.4

Figure 5 shows the probability of wrong exclusion (\mathcal{P}_{WE}). The over-identification cases (\mathcal{P}_{over+} and \mathcal{P}_{over-}) were smaller than 0.001 (i.e. they were practically null). There were not statistical overlap (\mathcal{P}_{ol}) for clusters 2, 3 and 4. We will discuss more about statistical overlap (\mathcal{P}_{ol}) later.

Fig 5. Probability of wrong exclusion (\mathcal{P}_{WE}) for the case of hard constraints and for $\alpha' = 0.001$. Cluster 1(a), Cluster 2(b), Cluster 3(c) and Cluster 4(d).

Result of the Soft Constraint Effects on the Iterative Outlier Elimination Procedure

The both configuration in Fig 3(d) and Figure 3(e) were analysed in terms of soft constraints. In that case, the critical values were $\hat{k} = 3.95$, $\hat{k} = 3.95$ and $\hat{k} = 3.92$ for two soft constraints with $\sigma_c = 0.1mm$, $\sigma_c = 1mm$ and $\sigma_c = 10mm$, respectively. In the case of three soft constraints, the critical values found were $\hat{k} = 3.99$, $\hat{k} = 3.99$ and $\hat{k} = 3.96$ for $\sigma_c = 0.1mm$, $\sigma_c = 1mm$ and $\sigma_c = 10mm$, respectively. All these critical

values were computed for $\alpha' = 0.001$. Table 3 gives the local redundancy (r_i), the standard-deviation of the LS-estimated outlier σ_{∇_i} and the maximum absolute correlation ($max_{\rho_{w_i, w_j}}$) for the scenarios of two constraints.

Table 3. Local redundancy (r_i), standard-deviation of the LS-estimated outlier σ_{∇_i} (mm) and the maximum absolute correlation ($max_{\rho_{w_i, w_j}}$) for each scenario of two soft constraints.

Measurement	$\sigma_c = 0.1mm$			$\sigma_c = 1mm$			$\sigma_c = 10mm$		
	r_i	σ_{∇_i}	$max_{\rho_{w_i, w_j}}$	r_i	σ_{∇_i}	$max_{\rho_{w_i, w_j}}$	r_i	σ_{∇_i}	$max_{\rho_{w_i, w_j}}$
y_1	0.581	1.312	0.564	0.471	1.457	0.681	0.397	1.587	0.994
y_2	0.582	1.311	0.376	0.533	1.369	0.423	0.501	1.413	0.471
y_3	0.581	1.312	0.564	0.471	1.457	0.681	0.397	1.587	0.994
y_4	0.581	1.312	0.564	0.471	1.457	0.681	0.397	1.587	0.994
y_5	0.582	1.311	0.376	0.533	1.369	0.423	0.501	1.413	0.471
y_6	0.581	1.312	0.564	0.471	1.457	0.681	0.397	1.587	0.994
y_7	0.583	1.310	0.359	0.571	1.324	0.423	0.563	1.333	0.471
y_8	0.583	1.310	0.359	0.571	1.324	0.423	0.563	1.333	0.471
y_9	0.583	1.310	0.359	0.571	1.324	0.423	0.563	1.333	0.471
y_{10}	0.583	1.310	0.359	0.571	1.324	0.423	0.563	1.333	0.471
y_{11}	0.583	1.309	0.358	0.583	1.309	0.398	0.583	1.309	0.433
y_{12}	0.583	1.309	0.358	0.583	1.309	0.398	0.583	1.309	0.433
y_{13}	0.007	1.163	1.000	0.300	1.826	1.000	0.497	14.189	1.000
y_{14}	0.007	1.163	1.000	0.300	1.826	1.000	0.497	14.189	1.000

From Table 3, five clusters were defined for each case of two soft constraints, i.e. for the case where heights A and D were given as soft constraints in Figure 3(d), as follows:

- Cluster 1: y_1, y_3, y_4 and y_6 .
- Cluster 2: y_2 and y_5 .
- Cluster 3: y_7, y_8, y_9 and y_{10} .
- Cluster 4: y_{11} and y_{12} .
- Cluster 5: y_{13} and y_{14} .

The probabilities of correct identification (\mathcal{P}_{CI}) and correct detection (\mathcal{P}_{CD}) for the measurements (Cluster 1 to Cluster 4) subject to the scenarios of two soft constraints (heights A and D) are displayed in Figure 6.

Fig 6. Probability of correct identification (\mathcal{P}_{CI}) and Probability of correct detection (\mathcal{P}_{CD}) for the measurements subject to the scenarios of two soft constraints for $\alpha' = 0.001$. Cluster 1(a,b), Cluster 2(c,d), Cluster 3(e,f) and Cluster 4(g,h).

Note that the Cluster 5 is associated with the two soft constraints (i.e. y_{13} and y_{14}). The probabilities of correct identification (\mathcal{P}_{CI}) for these both soft constraints were null. However, the probabilities of correct detection were not. Figure 7 shows the probabilities of correct detection (\mathcal{P}_{CD}) for these two soft constraints (i.e. heights A and D).

Fig 7. Probability of correct detection \mathcal{P}_{CD} for the two soft constraints and for $\alpha' = 0.001$. Cluster 5: heights A and D.

The probability of wrong exclusion (\mathcal{P}_{WE}) for the measurements (Cluster 1 to Cluster 4) subject to the scenarios of two soft constraints (heights A and D) are displayed in Figure 8. Figure 9 gives the the probability of wrong exclusion \mathcal{P}_{WE} for that two constraints (i.e. heights A and D). The over-identification cases (\mathcal{P}_{over+} and \mathcal{P}_{over-}) and the statistical overlap (\mathcal{P}_{ol}) were practically null for that case. The sensitivity indicators (MDB and MIB) for each scenario of two soft constraints are displayed in Table 4.

Fig 8. Probability of wrong exclusion \mathcal{P}_{WE} for the measurements subject to the scenarios of two soft constraints for $\alpha' = 0.001$. Cluster 1(a), Cluster 2(b), Cluster 3(c) and Cluster 4(d).

Fig 9. Probability of wrong exclusion \mathcal{P}_{WE} for the two soft constraints and for $\alpha' = 0.001$. Cluster 5: heights A and D.

Table 4. MDB and MIB for the case of two soft constraints based on $\alpha' = 0.001$ and $\tilde{\mathcal{P}}_{CD} = \tilde{\mathcal{P}}_{CI} = 0.8$.

Cluster	$\sigma_c = 10mm$		$\sigma_c = 1mm$		$\sigma_c = 0.1mm$	
	MDB (σ)	MIB (σ)	MDB (σ)	MIB (σ)	MDB (σ)	MIB (σ)
1	7.5	25	7	7.1	6.3	6.3
2	6.8	6.8	6.6	6.6	6.3	6.3
3	6.4	6.4	6.4	6.4	6.3	6.3
4	6.3	6.3	6.3	6.3	6.3	6.3
5	6.8	-	8.8	-	57	-

Table 5 gives the local redundancy (r_i), the standard-deviation of the LS-estimated outlier σ_{∇_i} and the maximum absolute correlation ($max_{\rho_{w_i, w_j}}$) for the scenarios of three soft constraints.

Table 5. Local redundancy (r_i), standard-deviation of the LS-estimated outlier σ_{∇_i} (mm) and the maximum absolute correlation ($max_{\rho_{w_i, w_j}}$) for each scenario of three soft constraints.

Measurement	$\sigma_c = 0.1mm$			$\sigma_c = 1mm$			$\sigma_c = 10mm$		
	r_i	σ_{∇_i}	$max_{\rho_{w_i, w_j}}$	r_i	σ_{∇_i}	$max_{\rho_{w_i, w_j}}$	r_i	σ_{∇_i}	$max_{\rho_{w_i, w_j}}$
y_1	0.702	1.194	0.660	0.502	1.411	0.577	0.398	1.586	0.992
y_2	0.582	1.311	0.326	0.533	1.369	0.412	0.501	1.413	0.470
y_3	0.702	1.194	0.660	0.502	1.411	0.577	0.398	1.586	0.992
y_4	0.702	1.194	0.660	0.502	1.411	0.577	0.398	1.586	0.992
y_5	0.582	1.311	0.326	0.533	1.369	0.412	0.501	1.413	0.470
y_6	0.702	1.194	0.660	0.502	1.411	0.577	0.398	1.586	0.992
y_7	0.704	1.192	0.415	0.602	1.289	0.412	0.563	1.333	0.470
y_8	0.704	1.192	0.415	0.602	1.289	0.412	0.563	1.333	0.470
y_9	0.704	1.192	0.415	0.602	1.289	0.412	0.563	1.333	0.470
y_{10}	0.704	1.192	0.415	0.602	1.289	0.412	0.563	1.333	0.470
y_{11}	0.583	1.309	0.326	0.583	1.309	0.385	0.583	1.309	0.433
y_{12}	0.583	1.309	0.326	0.583	1.309	0.385	0.583	1.309	0.433
y_{13}	0.012	0.904	0.660	0.425	1.534	0.542	0.663	12.283	0.501
y_{14}	0.012	0.904	0.660	0.425	1.534	0.542	0.663	12.283	0.501
y_{15}	0.019	0.718	0.63	0.5	1.414	0.542	0.665	12.268	0.501

The probabilities of correct identification (\mathcal{P}_{CI}) and correct detection (\mathcal{P}_{CD}) in Figure 10 were computed for the clusters based on Table 5, as follows:

- Cluster 1: y_1, y_3, y_4 and y_6 . 273
- Cluster 2: y_2 and y_5 . 274
- Cluster 3: y_7, y_8, y_9 and y_{10} . 275
- Cluster 4: y_{11} and y_{12} . 276
- Cluster 5: y_{13} and y_{14} . 277
- Cluster 6: y_{15} . 278

Fig 10. Probability of correct identification (\mathcal{P}_{CI}) and Probability of correct detection (\mathcal{P}_{CD}) for the measurements subject to the scenarios of three soft constraints for $\alpha' = 0.001$. Cluster 1(a,b), Cluster 2(c,d), Cluster 3(e,f) and Cluster 4(g,h).

Figure 11 shows the probabilities of correct identification \mathcal{P}_{CI} and correct detection \mathcal{P}_{CD} for the three soft constraints, i.e. for Cluster 5 (heights A and D) and Cluster 6 (height G) in Figure 3(e). The probability of wrong exclusion \mathcal{P}_{WE} for the measurements (Cluster 1 to Cluster 4) subject to the scenarios of three soft constraints (heights A, D and G) are displayed in Figure 12. Figure 13 gives the the probability of wrong exclusion \mathcal{P}_{WE} for that three constraints (i.e. heights A, D and G). The over-identification cases (\mathcal{P}_{over+} and \mathcal{P}_{over-}) and the statistical overlap (\mathcal{P}_{ol}) were also practically null for that case of three soft constraints. The sensitivity indicators (MDB and MIB) for each scenario of three soft constraints are displayed in Table 6. 279-287

Fig 11. Probability of correct identification (\mathcal{P}_{CI}) and Probability of correct detection (\mathcal{P}_{CD}) for the three constraints and for $\alpha' = 0.001$. Cluster 5(a,b) and Cluster 6(c,d).

Fig 12. Probability of wrong exclusion \mathcal{P}_{WE} for the measurements subject to the scenarios of three soft constraints and for $\alpha' = 0.001$. Cluster 1(a), Cluster 2(b), Cluster 3(c) and Cluster 4(d).

Fig 13. Probability of wrong exclusion (\mathcal{P}_{WE}) for the three constraints and for $\alpha' = 0.001$. Cluster 6(b).

Table 6. MDB and MIB for the case of three soft constraints based on $\alpha' = 0.001$ and $\tilde{\mathcal{P}}_{CD} = \tilde{\mathcal{P}}_{CI} = 0.8$.

Cluster	$\sigma_c = 10mm$		$\sigma_c = 1mm$		$\sigma_c = 0.1mm$	
	MDB (σ)	MIB (σ)	MDB (σ)	MIB (σ)	MDB (σ)	MIB (σ)
1	7.5	22	6.8	6.9	5.8	5.9
2	6.8	6.9	6.6	6.7	6.4	6.4
3	6.4	6.4	6.3	6.3	5.8	5.8
4	6.3	6.3	6.3	6.3	6.3	6.3
5	5.9	6.0	7.4	7.5	43.5	45
6	5.9	5.9	6.9	6.9	34.6	35.5

Discussion

We start from the scenario of one hard constraint in Figure 3(a). Table 1 reveal that the maximum correlation between w -test statistics for the measurements constituting Cluster 1 is exactly equal to 1.00 (i.e. $\max_{\rho_{w_i, w_j}} = 1.00$). This means that the measurements belonging to Cluster 1 are those that are connected with unknown heights whose connections are limited to only two. The both unknown heights A and D are tied only to two measurements (i.e. y_1 and y_6 linked to A, and y_3 and y_4 linked to D), therefore if an outlier occurred in one of these measurements, we would only be able to analyse the consistency between them, but we would not be able to distinguish which of them was contaminated by an outlier. This means that we would only be able to detect, because the w -test statistics could be larger than a critical value \hat{k} , however, in that case, the values of w -test statistics would be the same, and therefore we would not have only one unique maximum w -test statistics, but actually would have four maximum w -test statistics. In other words, the equation systems associated with the measurements of Cluster 1 are linearly dependent [65]. Therefore, there is no reliability in terms of outlier identification for the Cluster 1, as can be seen in Figure 3(a).

From Figure 3(b), we can note that there is reliability in terms of outlier detection for Cluster 1, and it is caused by overlapping w -test statistics. The probability of statistics overlap \mathcal{P}_{ol} for Cluster 1 in this scenario of minimally constrained network is displayed in Figure 14.

Fig 14. Probability of correct detection \mathcal{P}_{CD} and statistical overlap (\mathcal{P}_{ol}) for the Cluster 1 subject to one hard constraint and for $\alpha' = 0.001$.

Probability of correct detection \mathcal{P}_{CD} and statistical overlap (\mathcal{P}_{ol}) for the Cluster 1 subject to one hard constraint and for $\alpha' = 0.001$.

The problem of not having more connections (i.e. more measurements) for the unknown heights A and D in the case of one hard constraint with G fixed is overcome when these heights (A and D) are taken as hard constraints in 3(b) or when the heights A, D and G are hard constraints in figure 3(c). Figure 3(a, b) show that the measurements of Cluster 1 are now able to identify an outlier when two hard constraints (A and D fixed) are in play. It is also verified to the case of three hard constraints (A, D and G fixes) in Figure 3(e, f), i.e. there is reliability in terms of both outlier detection and identification for these measurements in those conditions.

From Table 2, we observe different behaviour for the clusters as follows:

- Cluster 1: there was no MIB for the case of having only one single hard constraint, whereas $MDB = MIB$ for the other cases. However, both MDB and MIB decrease significantly with the increase in the number of hard constraints.
- Cluster 2: MDB slightly smaller than MIB. Both MDB and MIB were practically the same for the case of having two or three hard constraints.
- Cluster 3: $MDB = MIB$ for all cases of hard constraints, however both MDB and MIB decrease significantly with the increase in the number of hard constraints.
- Cluster 4: MDB e MIB were equal for all cases.

In terms of outlier detection and identification, therefore, Cluster 1 was more sensitive to constraints, Cluster 3 relatively sensitive to constraints, whereas the Cluster 4 completely insensitive to constraints and Cluster 2 relatively insensitive to constraints. This is also can be seen in Figure 4. The reason for this is that the local redundancy (r_i) of the Cluster 1 increased with the increase of the number of hard constraints, whereas the Cluster 4 have remained the same, as can be seen in Table 2.

Leaving aside the cases of statistical overlap (\mathcal{P}_{ol}), the network presents low least-squares residuals correlation ($\rho_{w_i, w_j} < 0.5$) and high local redundancy ($r_i > 0.5$). Because of this, the probabilities of wrong exclusion (\mathcal{P}_{WE}) were less than 1%, as can be seen in Figure 5. The over-identification cases (\mathcal{P}_{over+} and \mathcal{P}_{over-}) were practically null. Consequently, $\mathcal{P}_{CI} \approx \mathcal{P}_{CD}$. Due of this fact, the family-wise error rate (α') should be increased in order to have more success rate in the outlier detection and identification [44].

From Figure 15, we observe that increasing the α' increases the both probabilities of correct identification (\mathcal{P}_{CI}) and correct detection (\mathcal{P}_{CD}) for outlier magnitude from 5σ to 6σ in the case of three hard constraints and from 5σ to 6.8σ in the case of two hard constraints. Although the rates of \mathcal{P}_{over+} and \mathcal{P}_{WE} also increase, they are not significant when compared to the improvement of correct identification (\mathcal{P}_{CI}) and detection (\mathcal{P}_{CD}). This same analysis can be done for the other clusters.

Fig 15. Probabilities of correct identification \mathcal{P}_{CI} , correct detection \mathcal{P}_{CD} , over-identification positive \mathcal{P}_{over+} and wrong exclusion \mathcal{P}_{WE} for the Cluster 1 subject to two and three hard constraints and for $\alpha' = 0.001$ and $\alpha' = 0.1$. Probabilities of correct identification \mathcal{P}_{CI} (a), correct detection \mathcal{P}_{CD} (b), over-identification positive \mathcal{P}_{over+} (c) and wrong exclusion \mathcal{P}_{WE} for the Cluster 1 subject to two and three hard constraints and for $\alpha' = 0.001$ and $\alpha' = 0.1$.

In terms of soft constraints for the cases of two constraints in Figure 3(d), we observe from Table 3 that the larger the relaxation of the constraint (i.e. the larger the standard-deviation of the constraint σ_c), the larger the residuals correlation (ρ_{w_i, w_j}) and the standard-deviation of the outlier σ_{∇_i} , and the smaller the local redundancy (r_i). Consequently, the probabilities of correct identification (\mathcal{P}_{CI}) and detection (\mathcal{P}_{CD}) get smaller and smaller with the relaxation of the constraints, whereas the probability of wrong exclusion gets larger (\mathcal{P}_{WE}). This can be more clearly verified in Figure 6(a, b) and Figure 8(a) for Cluster 1, whose measurements are connected with the constraints A and D (i.e. y_{13} and y_{14} in Table 3, respectively).

Note from Figure 8(a) that the probability of wrong exclusion (\mathcal{P}_{WE}) increases as the magnitude of the outlier (∇_i) increases. However, this is true up to a certain limit of outlier magnitude. The effect of residuals correlation ρ_{w_i, w_j} on the rates of wrong exclusion (\mathcal{P}_{WE}) and correct identification (\mathcal{P}_{CI}) tends to decrease with the increase in the magnitude of the outlier ∇_i . This effect is more clearly verified for Cluster 1 in case where the precision of the constraints are ten times worse than the measurements $\sigma_c = 10\sigma = 10mm$.

Note from Figure 6 that identifying an outlier in the Cluster 1 (i.e. y_1, y_3, y_4 and y_6) when $\sigma_c = 10mm$ is more difficult than the other clusters. This is due to the fact the Cluster 1 has a higher residuals correlation $\rho_{w_i, w_j} = 0.994$ than other clusters. We observe that the larger the relaxation of the constraints, the larger is the effect of the correlation ρ_{w_i, w_j} on the success rate of outlier identification (\mathcal{P}_{CI}). Consequently, the higher the sensitivity indicator for outlier identification (MIB). Table 2 reveals that the ratio between MIB and MDB for the Cluster 1 and for the scenario where the standard-deviations of that two soft constraints are $\sigma_c = 10mm$ is $MIB/MDB=25/7.5=3.3$. On the other hand, the relationship between MIB and MDB is practically one (i.e. $MIB/MDB=1.0$) for the others scenarios.

If the FWE rate (α') were increased for the case where the two soft constraints of $\sigma_c = 10mm$ are in play, we would not have great advantages for Cluster 1, due to its high residuals correlation ($\rho_{w_i, w_j} = 99.4\%$). From Figure 16, we can observe that the probability of correct identification (\mathcal{P}_{CI}) for outlier magnitudes from 5σ to 8σ is effectively larger to a user-defined $\alpha' = 0.1$ than $\alpha' = 0.001$. However, the success rate is still less than 80%, i.e. $\mathcal{P}_{CI} < 0.8$. Note, for example, the correct identification rate is

$\mathcal{P}_{CI} = 56\%$ for an outlier magnitude of $\nabla_i = 8\sigma$ and $\alpha' = 0.1$. For $\alpha' = 0.1$ the $MIB = 33.5\sigma = 33.5mm$, whereas for $\alpha' = 0.001$ is $MIB = 25\sigma = 25mm$. Therefore, in that case, the MIB for $\mathcal{P}_{CI} = 0.8(80\%)$ and $\alpha' = 0.1$ would be 34% larger than user-defined $\alpha' = 0.001$.

Fig 16. Probabilities of correct identification (\mathcal{P}_{CI}) for the Cluster 1 subject to two soft constraints (2 s.c.) A and D for $\alpha' = 0.001$ and $\alpha' = 0.1$. Probabilities of correct identification (\mathcal{P}_{CI}) for the Cluster 1 subject to two soft constraints (2 s.c.) A and D for $\alpha' = 0.001$ and $\alpha' = 0.1$.

The soft constraints A and D were grouped in the Cluster 5 (i.e. A and D were treated as pseudo-observations in the model). There is no reliability in terms of outlier identification for the constraints, because the residual correlation between them is $\rho_{w_i, w_j} = 100\%$, as can be seen in Table 3 for y_{13} and y_{14} . However, these soft constraints are able to detect an outlier. In that case, the probability of correct detection \mathcal{P}_{CD} in Figure 7 is mainly caused by the statistical overlap \mathcal{P}_{ol} , as can be seen in the example of the case of $\sigma_c = 10mm$ in Figure 17. From Table 4, we observe that the larger the relaxation of the constraints, the larger the MDB. Note that the values of MDB are given in σ , and therefore the MDB for $\sigma_c = 10mm$ is larger than $\sigma_c = 1mm$ and $\sigma_c = 0.1mm$, i.e. we had the following inequality:
 $MDB = 6.8\sigma_c = 6.8 \times 10mm = 68mm > MDB = 8.8\sigma_c = 8.8 \times 1mm = 8.8mm > MDB = 57\sigma_c = 57 \times 0.1mm = 5.7mm$. In that case, if the FWE (α') were increased, the rate of outlier detection by the Cluster 4 (i.e. by the soft constraints) would increase.

Fig 17. Probabilities of correct detection \mathcal{P}_{CD} and statistical overlap \mathcal{P}_{ol} for the two soft constraints A and D (Cluster 5) with $\sigma_c = 10mm$ and for $\alpha' = 0.001$. Probabilities of correct detection \mathcal{P}_{CD} and statistical overlap \mathcal{P}_{ol} for the two soft constraints A and D (Cluster 5) with $\sigma_c = 10mm$ and for $\alpha' = 0.001$.

Similar effects of the relaxation of the constraints on the performance of the *IDS* in case of two soft constraints are verified in case of three soft constraints, as can be seen in the Figures 10, 11, 12 and 13.

In case of having three soft constraints in Figure 3(e), there is reliability in terms of outlier identification for the three pseudo-observations y_{13} , y_{14} and y_{15} (i.e. for A, D and G), as can be seen in Figure 10 and Table 6. In that case, we also observe that the probabilities of correct detection \mathcal{P}_{CD} of the soft constraints A and D (i.e. Cluster 5) were approximately 13% for $\sigma_c = 10mm$, 16% for $\sigma_c = 1mm$ and 24% for $\sigma_c = 0.1mm$ larger than the scenario of the network subject to two soft constraints. Table 6 reveals that the advantage of having three soft constraints instead of two constraints is that the constraints become identifiable in the presence of an outlier. The behaviour of the probabilities of correct detection (\mathcal{P}_{CD}), correct identification (\mathcal{P}_{CI}) and wrong exclusion (\mathcal{P}_{WE}) was similar to the case of the two soft constraints. Furthermore, the larger the relaxation of the constraints, the smaller the residuals correlation between the measurements and the soft constraints and the larger the residuals correlation among the measurements.

We also observe that the case of two soft constraints for $\sigma_c = 0.1mm$ was comparable with two hard constraints (see e.g. Table 2 and Table 6) in terms of the probability levels associated with *IDS* for the measurements (i.e. clusters 1, 2, 3 and 4). In the same way for the case of two soft constraints with $\sigma_c = 1mm$ or $\sigma_c = 10mm$, the probabilities levels were similar to the one hard constraint for that measurements, with the benefit of two soft constraints having reliability in terms of outlier identification for the Cluster 1. Finally, the three soft constraints with $\sigma_c = 1mm$ and $\sigma_c = 10mm$ were comparable to the two soft constraints for that scenario of constraints relaxation,

wheres the three soft constraints for $\sigma_c = 0.1mm$ showed similar outcomes with three
hard constraints for the measurements (see e.g. Table 2 and Table 6). In that case,
however, an advantage of the three soft constraints on the three hard constraints is the
possibility of analysing the sensitivity of the constraints. It is important to emphasised
that the stochastic models of the measurements and constraints were assumed
well-known and defined for the analyzes performed here.

Conclusion

In general, the probability of correct identification (\mathcal{P}_{CI}) for the case of hard
constraints is larger than soft constraints. It can be verified in Figure 4, Figure 6 and
Figure 10. Therefore, hard constraints should be used in the stage of pre-processing
data for the purpose of identifying and removing possible outlying measurements. In
that process, one should opt to set out the redundant hard constraints at points in the
network where the smallest connections exist. After identifying and removing possible
outliers, the soft constraints should be employed to propagate the uncertainties of the
constraints (pseudobservations) to the model parameters during the process of
least-squares estimation. This recommendation is valid for outlier detection and
identification purpose.

We highlight the main findings of this research as follows:

- Under a system of a high local redundancy $r_i > 0.5$ and low residuals correlation ($\rho_{\mathbf{w}_i, \mathbf{w}_j} < 0.5$), if one increase the *family-wise error rate* (FWE) of the test statistic, the performance of the procedure will be improved for both scenarios of hard constraints and soft constraints.
- The larger the relaxation of the constraints, the larger is the effect of the residuals correlation ($\rho_{\mathbf{w}_i, \mathbf{w}_j}$) on the success rate of outlier identification (\mathcal{P}_{CI}). Consequently, the higher the sensitivity indicator for outlier identification (MIB), and therefore more difficult it becomes to identify an outlier.
- Under a scenario of soft constraints, one should set out at least three soft constraints in order to identify an outlier in the constraints.

In other types of analysis, for example, deformation analysis of geodetic networks one should formulated the constraints as pseudo-observations with $\sigma_c > 0$ in an adjustment model of system equations. In other words, the relaxation of constraints makes it possible to estimate deformation effects that are unmodelled in the deterministic model matrix, for example, deformations which have a spatial and temporal variations. In that case, the deformation can be modelled by covariance functions that determine the cofactors between constraints [20]. This should be applied to control points located in the structure, but not to control points located outside the structure. The control points located outside the structure should be hard constraints and have their stability verified by an independent process in order to analyse the deformation of the structure. So, one will be also able to analyse the sensitivity in terms of minimal detectable deformations and/or minimal identifiable deformations [14]. Future researches will be addressed to sensitivity analysis for deformation monitoring networks.

Supporting information

S1 Appendix. Description of the method. Provides a broad theoretical framework and detailed description of the method used to estimate the Iterative Data-Snooping probability levels.

Acknowledgments

The authors would like to thank CNPq—Conselho Nacional de Desenvolvimento Científico e Tecnológico—Brasil (proc. n^o 103587/2019-5) and PETROBRAS (Grant Number 2018/00545-0) for funding the research.

References

1. Fang X. Weighted total least-squares with constraints: a universal formula for geodetic symmetrical transformations. *Journal of Geodesy*. 2015;89(5):459–469. doi:10.1007/s00190-015-0790-8.
2. Amiri-Simkooei AR. Weighted Total Least Squares with Singular Covariance Matrices Subject to Weighted and Hard Constraints. *J Surv Eng*. 2017;143(4):04017018. doi:10.1061/(ASCE)SU.1943-5428.0000239.
3. Courant R, Hilbert D. *Methods of Mathematical Physics*. vol. 1. Wiley-VCH; 1989.
4. Grafarend EW. Optimization of Geodetic Networks. *The Canadian Surveyor*. 1974;28(5):716–723. doi:10.1139/tcs-1974-0120.
5. Teunissen P. Zero Order Design: Generalized Inverses, Adjustment, the Datum Problem and S-Transformations. In: Grafarend EW, Sansò F, editors. *Optimization and Design of Geodetic Networks*. Berlin, Heidelberg: Springer Berlin Heidelberg; 1985. p. 11–55.
6. Dermanis A. Free network solutions with the DLT method. *ISPRS Journal of Photogrammetry and Remote Sensing*. 1994;49(2):2 – 12. doi:https://doi.org/10.1016/0924-2716(94)90061-2.
7. Kotsakis C. Reference frame stability and nonlinear distortion in minimum-constrained network adjustment. *Journal of Geodesy*. 2012;86(9):755–774. doi:10.1007/s00190-012-0555-6.
8. Kotsakis C. In: Grafarend E, editor. *Datum Definition and Minimal Constraints*. Cham: Springer International Publishing; 2018. p. 1–6. Available from: https://doi.org/10.1007/978-3-319-02370-0_157-1.
9. Matsuoka MT, Rofatto VF, Klein I, Roberto Veronez M, da Silveira LG, Neto JBS, et al. Control Points Selection Based on Maximum External Reliability for Designing Geodetic Networks. *Applied Sciences*. 2020;10(2). doi:10.3390/app10020687.
10. Baarda W. S-transformations and criterion matrices. *Publ on geodesy, New Series*. 1973;5(2).
11. Schaffrin B. Aspects of Network Design. In: Grafarend EW, Sansò F, editors. *Optimization and Design of Geodetic Networks*. Berlin, Heidelberg: Springer Berlin Heidelberg; 1985. p. 548–597.
12. XU P. A general solution in geodetic nonlinear rank-defect models. *Bollettino di geodesia e scienze affini*. 1997;.

13. Altamimi Z, Dermanis A. The Choice of Reference System in ITRF Formulation. In: Sneeuw N, Novák P, Crespi M, Sansò F, editors. VII Hotine-Marussi Symposium on Mathematical Geodesy. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 329–334.
14. Velsink H. On the deformation analysis of point fields. *J Geod.* 2015;89(11):1071–1087. doi:10.1007/s00190-015-0835-z.
15. Velsink H. Extendable linearised adjustment model for deformation analysis. *Survey Review.* 2015;47(345):397–410. doi:10.1179/1752270614Y.0000000140.
16. Velsink H. Time Series Analysis of 3D Coordinates Using Nonstochastic Observations. *Journal of Applied Geodesy.* 2016;10(1):5–16.
17. Rao CR. Markoff's Theorem with Linear Restrictions on Parameters. *Sankhyā: The Indian Journal of Statistics (1933-1960).* 1945;7(1):16–19.
18. Lehmann R, Neitzel F. Testing the compatibility of constraints for parameters of a geodetic adjustment model. *Journal of Geodesy.* 2013;87(6):555–566. doi:10.1007/s00190-013-0627-2.
19. Velsink H. Testing Methods for Adjustment Models with Constraints. *Journal of Surveying Engineering.* 2018;144(4):04018009. doi:10.1061/(ASCE)SU.1943-5428.0000260.
20. Velsink H. Testing deformation hypotheses by constraints on a time series of geodetic observations. *J Appl Geod.* 2017;12(1):77–93. doi:10.1515/jag-2017-0028.
21. Baarda W. A testing procedure for use in geodetic networks. *Publ on geodesy, New Series.* 1968;2(5).
22. Teunissen PJG. First and second moments of non-linear least-squares estimators. *Bull Geodesique (Journal of Geodesy).* 1989;63:253–262. doi:https://doi.org/10.1007/BF02520475.
23. Teunissen PJG. *Testing Theory: an introduction.* 2nd ed. Delft University Press; 2006.
24. Amiri-Simkooei AR. *Least-squares variance component estimation: theory and GPS applications [PhD thesis].* Delft University of Technology; 2007. Available from: <https://ncgeo.nl/index.php/en/publicatiesgb/publications-on-geodesy>.
25. Koch KR. *Parameter estimation and hypothesis testing in linear models.* 2nd ed. Springer; 1999.
26. Tao H, Feng H, Xu L, Miao M, Long H, Yue J, et al. Estimation of Crop Growth Parameters Using UAV-Based Hyperspectral Remote Sensing Data. *Sensors.* 2020;20(5). doi:10.3390/s20051296.
27. Han M, Wang Q, Wen Y, He M, He X. The Application of Robust Least Squares Method in Frequency Lock Loop Fusion for Global Navigation Satellite System Receivers. *Sensors.* 2020;20(4). doi:10.3390/s20041224.
28. Wei C. Estimation for the Discretely Observed Cox-Ingersoll-Ross Model Driven by Small Symmetrical Stable Noises. *Symmetry.* 2020;12(3). doi:10.3390/sym12030327.

29. Sakic P, Ballu V, Royer JY. A Multi-Observation Least-Squares Inversion for GNSS-Acoustic Seafloor Positioning. *Remote Sensing*. 2020;12(3). doi:10.3390/rs12030448.
30. Livadiotis G. General Fitting Methods Based on Lq Norms and their Optimization. *Stats*. 2020;3(1):16–31. doi:10.3390/stats3010002.
31. Farooq SZ, Yang D, Ada ENJ. A Cycle Slip Detection Framework for Reliable Single Frequency RTK Positioning. *Sensors*. 2020;20(1). doi:10.3390/s20010304.
32. Araveeporn A. Comparing Parameter Estimation of Random Coefficient Autoregressive Model by Frequentist Method. *Mathematics*. 2020;8(1). doi:10.3390/math8010062.
33. Zhang C, Peng T, Zhou J, Ji J, Wang X. An Improved Autoencoder and Partial Least Squares Regression-Based Extreme Learning Machine Model for Pump Turbine Characteristics. *Applied Sciences*. 2019;9(19). doi:10.3390/app9193987.
34. Ji J, Yang M, Jiang L, He J, Teng Z, Liu Y, et al. Output-Only Parameters Identification of Earthquake-Excited Building Structures with Least Squares and Input Modification Process. *Applied Sciences*. 2019;9(4). doi:10.3390/app9040696.
35. Büchele D, Chao M, Ostermann M, Leenen M, Bald I. Multivariate chemometrics as a key tool for prediction of K and Fe in a diverse German agricultural soil-set using EDXRF. *Scientific Reports*. 2019;9(1):17588. doi:10.1038/s41598-019-53426-5.
36. Zhang J, Richardson JD, Dunkley BT. Classifying post-traumatic stress disorder using the magnetoencephalographic connectome and machine learning. *Scientific Reports*. 2020;10(1):5937. doi:10.1038/s41598-020-62713-5.
37. Yalage Don SM, Schmidtke LM, Gambetta JM, Steel CC. *Aureobasidium pullulans* volatilome identified by a novel, quantitative approach employing SPME-GC-MS, suppressed *Botrytis cinerea* and *Alternaria alternata* in vitro. *Scientific Reports*. 2020;10(1):4498. doi:10.1038/s41598-020-61471-8.
38. Bica R, Palarea-Albaladejo J, Kew W, Uhrin D, Pacheco D, Macrae A, et al. Nuclear Magnetic Resonance to Detect Rumen Metabolites Associated with Enteric Methane Emissions from Beef Cattle. *Scientific Reports*. 2020;10(1):5578. doi:10.1038/s41598-020-62485-y.
39. Chen X, Qiao W, Miao W, Zhang Y, Mu X, Wang J. The Dependence of Implicit Solvent Model Parameters and Electronic Absorption Spectra and Photoinduced Charge Transfer. *Scientific Reports*. 2020;10(1):3713. doi:10.1038/s41598-020-60757-1.
40. Weaving D, Jones B, Ireton M, Whitehead S, Till K, Beggs CB. Overcoming the problem of multicollinearity in sports performance data: A novel application of partial least squares correlation analysis. *PLOS ONE*. 2019;14(2):1–16. doi:10.1371/journal.pone.0211776.
41. Chen Y. Spatial Autocorrelation Approaches to Testing Residuals from Least Squares Regression. *PLOS ONE*. 2016;11(1):1–19. doi:10.1371/journal.pone.0146865.
42. Kargoll B. On the theory and application of model misspecification tests in geodesy [Doctoral thesis]. University of Bonn, Landwirtschaftliche Fakultät. German, Bonn; 2007.

43. Lehmann R. On the formulation of the alternative hypothesis for geodetic outlier detection. *J Geod.* 2013;87(4):373–386. doi:10.1007/s00190-012-0607-y.
44. Rofatto VF, Matsuoka MT, Klein I, Roberto Veronez M, da Silveira LG. A Monte Carlo-Based Outlier Diagnosis Method for Sensitivity Analysis. *Remote Sensing.* 2020;12(5). doi:10.3390/rs12050860.
45. Goldstein M, Uchida S. A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data. *PLOS ONE.* 2016;11(4):1–31. doi:10.1371/journal.pone.0152173.
46. Faria B, Vistulo de Abreu F. Cellular frustration algorithms for anomaly detection applications. *PLOS ONE.* 2019;14(7):1–31. doi:10.1371/journal.pone.0218930.
47. Aljably R, Tian Y, Al-Rodhaan M, Al-Dhelaan A. Anomaly detection over differential preserved privacy in online social networks. *PLOS ONE.* 2019;14(4):1–20. doi:10.1371/journal.pone.0215856.
48. El Azami M, Hammers A, Jung J, Costes N, Bouet R, Lartizien C. Detection of Lesions Underlying Intractable Epilepsy on T1-Weighted MRI as an Outlier Detection Problem. *PLOS ONE.* 2016;11(9):1–21. doi:10.1371/journal.pone.0161498.
49. Gautier M, Hocking TD, Foulley JL. A Bayesian Outlier Criterion to Detect SNPs under Selection in Large Data Sets. *PLOS ONE.* 2010;5(8):1–16. doi:10.1371/journal.pone.0011913.
50. George NI, Bowyer JF, Crabtree NM, Chang CW. An Iterative Leave-One-Out Approach to Outlier Detection in RNA-Seq Data. *PLOS ONE.* 2015;10(6):1–10. doi:10.1371/journal.pone.0125224.
51. Lehmann R. Improved critical values for extreme normalized and studentized residuals in Gauss–Markov models. *J Geod.* 2012;86(12):1137–1146. doi:10.1007/s00190-012-0569-0.
52. Rofatto VF, Matsuoka MT, Klein I, Veronez MR, Bonimani ML, Lehmann R. A half-century of Baarda’s concept of reliability: a review, new perspectives, and applications. *Surv Rev.* 2018;0(0):1–17. doi:10.1080/00396265.2018.1548118.
53. Zaminpardaz S, Teunissen PJG. DIA-datasnooping and identifiability. *J Geod.* 2019;93(1):85–101. doi:10.1007/s00190-018-1141-3.
54. Kok JJ, States U. On data snooping and multiple outlier testing [microform] / Johan J. Kok. U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Charting and Geodetic Services : For sale by the National Geodetic Information Center, NOAA Rockville, Md; 1984.
55. Knight NL, Wang J, Rizos C. Generalised measures of reliability for multiple outliers. *Journal of Geodesy.* 2010;84(10):625–635. doi:10.1007/s00190-010-0392-4.
56. Gui Q, Li X, Gong Y, Li B, Li G. A Bayesian unmasking method for locating multiple gross errors based on posterior probabilities of classification variables. *J Geod.* 2011;85(4):191–203. doi:10.1007/s00190-010-0429-8.
57. Klein I, Matsuoka MT, Guzzatto MP, Nievinski FG. An approach to identify multiple outliers based on sequential likelihood ratio tests. *Surv Rev.* 2017;49(357):449–457. doi:10.1080/00396265.2016.1212970.

58. Hawkins DM. Identification of Outliers. 1st ed. Springer Netherlands; 1980.
59. Algarni DA, Ali AE. Heighting and Distance Accuracy with Electronic Digital Levels. *Journal of King Saud University - Engineering Sciences*. 1998;10(2):229 – 239. doi:[https://doi.org/10.1016/S1018-3639\(18\)30698-6](https://doi.org/10.1016/S1018-3639(18)30698-6).
60. Takalo M, Rouhiainen P. Development of a System Calibration Comparator for Digital Levels in Finland. *Nordic Journal of Surveying and Real Estate Research*. 1;1(2).
61. Wiedemann W, Wagner A, Wunderlich T. Using IATS to Read and Analyze Digital Levelling Staffs. In: Paar R, Marendić A, Zrinjski M, editors. SIG 2016. Varazdin, Croatia: Croatian Geodetic Society; 2016. p. 515–526. Available from: http://www.geof.unizg.hr/pluginfile.php/7437/mod_book/chapter/173/TS6_2.pdf.
62. P J G T. The Geometry of Geodetic Inverse Linear Mapping and Non-linear Adjustment. *Publications on Geodesy, New Series*. 1985;8(1).
63. Rao CR, Mitra SK. Generalized inverse of a matrix and its applications; 1972.
64. Imparato D, Teunissen PJG, Tiberius CCJM. Minimal Detectable and Identifiable Biases for quality control. *Surv Rev*. 2019;51(367):289–299. doi:10.1080/00396265.2018.1437947.
65. Hekimoglu S, Erenoglu RC, Sanli DU, Erdogan B. Detecting Configuration Weaknesses in Geodetic Networks. *Survey Review*. 2011;43(323):713–730. doi:10.1179/003962611X13117748892632.

5 Considerações Finais

Neste trabalho, o principal objetivo foi estender a teoria da confiabilidade, originalmente proposta por Baarda (1968), para o controle de qualidade em geodésia, finalidade esta alcançada com êxito.

Aqui, nós fornecemos uma nova definição para o termo outlier: “outlier é uma observação que se afastou de seu valor mais provável a ponto de não pertencer ao modelo matemático (funcional e estocástico) estipulado”.

Yang et al. (2013) apresentaram uma teoria que estabelece os limites dos erros de decisão do Data-Snooping sob múltiplas hipóteses alternativas. Com essa teoria, as probabilidades de cometer erros do tipo I, II e III podem ser controladas para que a probabilidade de identificação de um outlier possa ser garantida ao executar o Data-Snooping. Eles estenderam a teoria convencional de Baarda, porém com algumas limitações: (i) fixaram a correlação entre as estatísticas de teste no estudo sobre a relação entre o número de observações e as probabilidades de cometer os tipos de erro (I, II e III); (ii) consideraram um único processo de estimação e teste, sem qualquer diagnóstico subsequente, como a remoção do outlier. Portanto, a confiabilidade foi estendida para o caso em que o Data-Snooping é executado uma única vez e para um valor fixo de correlação entre as estatísticas de teste.

Na prática, entretanto, temos: (i) resíduos com diferentes correlações entre si; (ii) Data-Snooping aplicado de forma iterativa: após a identificação e a eliminação de um possível outlier, os dados são reprocessados e a identificação é reiniciada. Este procedimento é denominado de Data-Snooping Iterativo (DSI). O DSI é, portanto, um caso que envolve não somente múltiplas hipóteses alternativas, mas também múltiplas rodadas de estimação, teste e adaptação. Logo, a probabilidade da correta identificação do outlier \mathcal{P}_{CI} não depende apenas dos erros Tipo I, Tipo II (\mathcal{P}_{MD}) e Tipo III (\mathcal{P}_{WE}), mas também das probabilidades de sobre-identificação (\mathcal{P}_{over+} e \mathcal{P}_{over-}) e sobreposição estatística (\mathcal{P}_{ol}).

Destacamos que as funções densidade de probabilidade associadas ao DSI são intratáveis por métodos determinísticos. Como um modelo analítico de fácil solução é desconhecido para o DSI, recorreremos ao método de Monte Carlo. A ideia central do método de Monte Carlo aqui foi repetir sucessivas amostragens aleatórias, um elevado número de vezes, para calcular as probabilidades heurísticamente, tal como se, de fato, se registrassem os resultados em experimentos reais. As amostragens aleatórias foram artificialmente geradas por meio de um gerador de números aleatórios.

O procedimento computacional aqui proposto foi desenvolvido para estimar os níveis de probabilidade do DSI. Com base nas estimativas da taxa de sucesso do DSI, o menor

outlier detectável (MDB – Minimal Detectable Bias) e o menor outlier identificável (MIB – Minimal Identifiable Bias) de cada observação podem ser obtidos para um certo nível de probabilidade estipulado. O MIB está associado com a confiabilidade interna em termos de identificação, e sua propagação para os parâmetros incógnitos do modelo, quando não identificado e eliminado corretamente, descreve a confiabilidade externa. O procedimento não faz uso de medidas reais coletadas em *situ* e, portanto, pode ser empregado na etapa de pré-análise ou projeto.

Importante destacar que o DSI é diferente do método DIA (*Detecção, Identificação e Adaptação*) (TEUNISSEN, 2018; ZAMINPARDAZ; TEUNISSEN, 2019a). No caso do DIA, a identificação só é iniciada quando a hipótese nula é rejeitada pelo teste global do modelo (*overall model test*). No DSI, por outro lado, o detector é o máximo valor absoluto da estatística de teste w_i . O teste global depende da redundância do sistema (i.e. $r = n - u$). Neste caso, como apontado por Klein (2011), quanto maior é $r = n - u$, maior é o valor crítico e, portanto, maior é tolerância do teste global do modelo a possíveis outliers. Por esse motivo, o procedimento DSI sempre deve ser realizado, pois a sua estatística de teste independe de $r = n - u$, ou seja, o DSI não se torna insensível a presença de outliers a medida que o número de graus de liberdade do modelo aumenta. Entretanto, como foi abordado, o DSI depende da redundância local e da correlação entre as estatísticas de teste.

Embora as estatísticas de teste para detecção e identificação de outliers existam há mais de oito décadas, a extensão da teoria da confiabilidade somente foi possível graças à disponibilidade de microcomputadores de baixo-custo, com poder computacional suficiente para descrever os níveis de probabilidade do DSI.

5.1 Principais Conclusões

As principais conclusões são destacadas abaixo para cada estudo realizado:

- O uso do valor crítico calculado por Monte Carlo é sempre melhor para fins de detecção de outlier, ou seja, $\mathcal{P}_{CD(\hat{k})} > \mathcal{P}_{CD(bon.f)}$.
- Para um sistema com baixa correlação entre os resíduos ($\rho_{w_i, w_j} < 0.5$) e alta redundância local $r_i > 0.5$, o uso do valor crítico calculado por Monte Carlo é melhor do que aquele calculado via Bonferroni, ou seja, $\mathcal{P}_{CI(\hat{k})} > \mathcal{P}_{CI(bon.f)}$ para $\rho_{w_i, w_j} < 0.5$ e $r_i > 0.5$.
- Para um sistema com alta correlação entre os resíduos ($\rho_{w_i, w_j} > 0.7$) e para o efeito de um possível outlier com magnitude próxima da incerteza do sistema, o valor crítico calculado por Monte Carlo é melhor do que Bonferroni, ou seja, $\mathcal{P}_{CI(\hat{k})} > \mathcal{P}_{CI(bon.f)}$.

- O efeito de um possível outlier ∇_i torna-se identificável quando há um decaimento simultâneo da contribuição individual de cada observação identificada erroneamente (\mathcal{P}_{WE}). Dessa forma, é possível encontrar o menor outlier identificável (Minimal Identifiable Bias – MIB) para um dado α' .
- Se for aplicado o DSI para o controle de qualidade dos dados observados, recomenda-se que as injunções absolutas sejam aplicadas no estágio de pré-processamento dos dados com o objetivo de identificar e remover possíveis outliers. Nesse processo, deve-se optar por restrições redundantes. No caso de redes geodésicas, as restrições devem ser localizadas nos pontos com menores graus de liberdade local. Após a remoção de possíveis outliers, o modelo poderá ser reconfigurado com suas respectivas injunções relativas (pseudo-observações) com o objetivo de propagar as suas incertezas para os parâmetros do modelo. Essa recomendação é válida apenas quando o objetivo é detectar e identificar anomalias que possam comprometer o modelo padrão.
- Se o objetivo for proteger os parâmetros do modelo de efeitos de possíveis outliers nos dados observados e, caso o DSI não esteja disponível e seja utilizado apenas o estimador de mínimos quadrados no processamento dos dados, é recomendável que sejam adotadas as injunções relativas para absorver os efeitos de possíveis outliers. Os resultados mostram que, quando o desvio-padrão das restrições são maiores do que o das observações ($\sigma_c > \sigma_i$), os parâmetros ficam menos suscetíveis a outliers, pois a redundância local r_i diminui com o aumento de (σ_c). Por esta razão, identificar um outlier torna-se mais difícil com a injunções relativas, pois ela absorve grande parte dos resíduos.
- Geralmente, alguns fenômenos de interesse são parametrizados no modelo e deseja-se que tais parâmetros absorvam parte dos resíduos. Nesse caso, as pseudo-observações (injunções relativas) são empregadas de modo que os resíduos estimados sejam absorvidos nessas restrições. Um exemplo prático das pseudo-observações é a sua aplicação na estimativa do atraso zenital troposférico nos sinais GNSS (Global Navigation Satellite System), em que o ruído da pseudo-observação, ou seja, sua incerteza, é intencionalmente aumentada com o objetivo de modelar os efeitos da troposfera. Quanto maiores forem as incertezas das pseudo-observações, maior será a variação estocástica e, portanto, maior será a taxa de detecção e identificação.

5.2 Perspectivas Futuras

Demonstramos que o DSI não é somente um procedimento de teste, mas um estimador, cuja finalidade é minimizar os efeitos dos outliers na estimativa dos parâmetros incógnitos do modelo matemático. As estimativas dependem dos níveis de probabilidade do DSI, ou seja, da probabilidade de correta detecção \mathcal{P}_{CD} e \mathcal{P}_{CI} .

Devido à sua estrutura, o estimador DSI não é linear. Consequentemente, o estimador DSI dificilmente terá distribuição normal. Investigações futuras deverão ser direcionadas com o objetivo de analisar a distribuição do estimador DSI. No que diz respeito ao estimador DIA, [Zaminpardaz e Teunissen \(2019b\)](#) faz importantes considerações, como segue: (i) Quanto menor o Erro Tipo I ou II, mais o estimador DIA se aproxima da normal; (ii) Quanto menor o Erro tipo II, menos enviesado é o estimador DIA; (iii) Quanto mais precisas as observações, mais achatada é a f.d.p. do estimador DIA; e (iv) Quanto menor o Erro tipo II, menos enviesado se torna o estimador DIA na hipótese alternativa. Essas considerações deverão ser ampliadas para o caso do DSI.

Alguns pontos que não foram abordados na tese serão alvos de pesquisas futuras, tais como:

- Desenvolver um estimador em que a função indicadora não seja apenas binária, ou seja, “0” para não identificação e “1” para identificação, mas de tal forma que as probabilidades de cada hipótese sejam levadas em consideração.
- Analisar os métodos que controlam o k -FWER, que é uma versão generalizada do FWER usual. É uma taxa de erro entre famílias de testes que tem sido introduzida recentemente da literatura estatística.
- Analisar a taxa de descoberta falsa, conhecida como False Discovery Rate (FDR), que é o valor esperado da razão entre o número de falsos positivos e o número total de testes rejeitados. Um procedimento que controla o FDR, mas não o FWER, parece ser mais poderoso, segundo a literatura estatística.
- Comparar a detecção do teste global do ajustamento e a detecção do teste local. Em outras palavras, comparar o DIA-datasnooping de [Zaminpardaz e Teunissen \(2019a\)](#) com o DSI.
- Integrar os métodos SLRTMO de [Klein et al. \(2017\)](#), DSI e p -value de [Lehmann e Lösler \(2016\)](#) no cenários de múltiplos outliers simultâneos.
- Usufruir do método de Monte Carlo para calcular incertezas de problemas de ordem complexa, por exemplo, estimar as propriedades de certos estimadores robustos.
- Avaliar o DSI no caso de múltiplos outliers simultâneos.
- Investigar o uso de simulações estocásticas em problemas de monitoramento de estruturas.
- Integrar o Método Monte Carlo com algoritmos de *machine learning* e *deep learning*, tais como redes neurais artificiais e meta-heurísticas.

Referências

- ABDI, H. The Bonferonni and Šidák Corrections for Multiple Comparisons. In: . [S.l.: s.n.], 2006. Citado na página 42.
- ALKHATIB, H.; NEUMANN, I.; KUTTERER, H. Uncertainty modeling of random and systematic errors by means of Monte Carlo and fuzzy techniques. *Journal of Applied Geodesy*, De Gruyter, Berlin, Boston, v. 3, n. 2, p. 67–79, 2009. Disponível em: <<https://www.degruyter.com/view/journals/jag/3/2/article-p67.xml>>. Citado na página 55.
- AMIRI-SIMKOOEI, A. *Least-Squares Variance Component Estimation: Theory and GPS Applications*. Tese (Doutorado) — Aerospace Engineering, TU Delft, 2007. Citado na página 164.
- AMIRI-SIMKOOEI, A. R. Weighted total least squares with singular covariance matrices subject to weighted and hard constraints. *J. Surv. Eng.*, v. 143, n. 4, p. 04017018, 2017. Disponível em: <<https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29SU.1943-5428.0000239>>. Citado na página 164.
- ARNOLD, S. *The theory of linear models and multivariate analysis*. 1. ed. [S.l.]: Wiley, 1981. Citado na página 25.
- AYDIN, C.; DEMIREL, H. Computation of Baarda's Lower Bound of the Non-Centrality Parameter. *J. Geod.*, v. 78, n. 7, p. 437–441, Mar 2004. ISSN 1432-1394. Disponível em: <<https://doi.org/10.1007/s00190-004-0406-1>>. Citado 2 vezes nas páginas 39 e 69.
- BAARDA, W. Statistical concepts in geodesy. *Publ. on geodesy, New Series*, v. 2, n. 4, 1967. Citado 2 vezes nas páginas 25 e 69.
- BAARDA, W. A testing procedure for use in geodetic networks. *Publ. on geodesy, New Series*, v. 2, n. 5, 1968. Citado 15 vezes nas páginas 8, 10, 24, 25, 28, 31, 32, 35, 36, 38, 39, 68, 69, 197 e 223.
- BASELGA, S. Second order design of geodetic networks by the simulated annealing method. *Journal of Surveying Engineering*, v. 137, n. 4, p. 167–173, 2011. Citado na página 56.
- BOX, G. E. P.; MULLER, M. E. A note on the generation of random normal deviates. *The Annals of Mathematical Statistics*, v. 29, n. 2, p. 610–611, 1958. <<https://doi:10.1214/aoms/1177706645>>. Citado na página 58.
- CHATTERJEE, S. et al. Monte Carlo simulation of globular cluster evolution v. binary stellar evolution. *The Astrophysical Journal*, IOP Publishing, v. 719, n. 1, p. 915–930, 2010. Citado na página 54.
- DU, Y.; GENG, Y.; SUN, L. Simulation model based on Monte Carlo method for traffic assignment in local area road network. *Frontiers of Architecture and Civil Engineering in China*, v. 3, n. 2, p. 195–203, Jun 2009. ISSN 1673-7512. Disponível em: <<https://doi.org/10.1007/s11709-009-0032-3>>. Citado na página 54.

- DUCHNOWSKI, R. Hodges–Lehmann estimates in Deformation Analyses. *Journal of Geodesy*, v. 87, n. 10, p. 873–884, Nov 2013. ISSN 1432-1394. Disponível em: <<https://doi.org/10.1007/s00190-013-0651-2>>. Citado na página 25.
- 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*. Deutsche Geodätische Kommission, Munich, Germany: [s.n.], 1983. Series A, n. 98, p. 79–103. ISSN 978376681802. Citado 3 vezes nas páginas 25, 43 e 69.
- GEMAEL, C. *Introdução ao ajustamento de observações: aplicações geodésicas*. 1. ed. [S.l.]: Editora da UFPR, 1994. Citado na página 24.
- GHILANI, C. D. *Adjustment Computations: Spatial Data Analysis*. 6. ed. [S.l.]: John Wiley Sons, Ltd, 2017. ISBN 9781119390664. Citado 3 vezes nas páginas 24, 26 e 28.
- HAWKINS, D. M. *Identification of Outliers*. 1. ed. [S.l.]: Springer Netherlands, 1980. <<https://doi.org/10.1007/978-94-015-3994-4>>. ISBN 9040719756. Citado 2 vezes nas páginas 24 e 25.
- HEWITSON, S.; WANG, J. GNSS receiver autonomous integrity monitoring (RAIM) performance analysis. *GPS Solutions*, v. 10, n. 3, p. 155–170, Jul 2006. ISSN 1521-1886. Disponível em: <<https://doi.org/10.1007/s10291-005-0016-2>>. Citado na página 43.
- HROMKOVIČ, J. *Algorithmics for Hard Problems: Introduction to Combinatorial Optimization, Randomization, Approximation, and Heuristics*. Springer-Verlag Berlin Heidelberg, 2001. Disponível em: <<https://www.springer.com/gp/book/9783662046166>>. Citado na página 54.
- HSU, J. *Multiple Comparisons*. 1st. ed. [S.l.]: Chapman and Hall/CRC, 1996. Citado na página 40.
- HUBER, P. *Robust Statistics*. [S.l.]: Wiley, 1981. (Wiley Series in Probability and Statistics). ISBN 9780471418054. Citado na página 25.
- IMPARATO, D.; TEUNISSEN, P.; TIBERIUS, C. Minimal detectable and identifiable biases for quality control. *Surv. Rev.*, Taylor Francis, v. 51, n. 367, p. 289–299, 2019. Disponível em: <<https://doi.org/10.1080/00396265.2018.1437947>>. Citado 3 vezes nas páginas 45, 55 e 69.
- JOINT COMMITTEE FOR GUIDES ON METROLOGY. *The international vocabulary of metrology—basic and general concepts and associated terms (VIM)*. 3. ed. [S.l.], 2012. Disponível em: <https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf>. Acesso em: 11 mar. 2020. Citado na página 67.
- JWO, D.-J.; CHEN, J.-J. GPS/INS Navigation Filter Designs Using Neural Network with Optimization Techniques. In: JIAO, L. et al. (Ed.). *Advances in Natural Computation*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. p. 461–470. ISBN 978-3-540-45902-6. Citado na página 56.
- KAVOURAS, M. *On the Detection of Outliers and the Determination of Reliability in Geodetic Networks*. Dissertação (Mestrado) — Department of Geodesy and Geomatics Engineering, University of New Brunswick, 1982. Citado na página 120.

- KIM, W. et al. Monte Carlo analysis of the accelerator-driven system at kyoto university research reactor institute. *Nuclear Engineering and Technology*, v. 48, n. 2, p. 304 – 317, 2016. ISSN 1738-5733. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1738573315002569>>. Citado na página 54.
- KLEIN, I. *Controle de Qualidade no Ajustamento de Observações Geodésicas*. Dissertação (Mestrado) — Programa de Pós-Graduação em Sensoriamento Remoto, UFRGS, 2011. Citado 4 vezes nas páginas 24, 26, 48 e 198.
- KLEIN, I. *Proposta de um novo método para o planejamento de redes geodésicas*. Tese (Doutorado) — Programa de Pós-Graduação em Sensoriamento Remoto, UFRGS, 2014. Citado na página 120.
- KLEIN, I. et al. On evaluation of different methods for quality control of correlated observations. *Surv. Rev.*, Taylor Francis, v. 47, n. 340, p. 28–35, 2015. Disponível em: <<https://doi.org/10.1179/1752270614Y.0000000089>>. Citado 2 vezes nas páginas 25 e 55.
- KLEIN, I. et al. An approach to identify multiple outliers based on sequential likelihood ratio tests. *Surv. Rev.*, Taylor Francis, v. 49, n. 357, p. 449–457, 2017. Disponível em: <<https://doi.org/10.1080/00396265.2016.1212970>>. Citado na página 200.
- KLEIN, I. et al. A new relationship between the quality criteria for geodetic networks. *Journal of Geodesy*, v. 93, n. 4, p. 529–544, Apr 2019. ISSN 1432-1394. Disponível em: <<https://doi.org/10.1007/s00190-018-1181-8>>. Citado na página 55.
- KNIGHT, N. L.; WANG, J.; RIZOS, C. Generalised measures of reliability for multiple outliers. *Journal of Geodesy*, v. 84, n. 10, p. 625–635, Oct 2010. ISSN 1432-1394. Disponível em: <<https://doi.org/10.1007/s00190-010-0392-4>>. Citado 2 vezes nas páginas 69 e 120.
- KOCH, I. et al. Robust estimators in geodetic networks based on a new metaheuristic: Independent vortices search. *Sensors*, v. 19, n. 20, 2019. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/19/20/4535>>. Citado na página 56.
- KOCH, K. R. *Parameter estimation and hypothesis testing in linear models*. 2. ed. [S.l.]: Springer, 1999. Citado na página 24.
- KOCH, K.-R. Evaluation of uncertainties in measurements by Monte Carlo simulations with an application for laserscanning. *Journal of Applied Geodesy*, De Gruyter, Berlin, Boston, v. 2, n. 2, p. 67 – 77, 2008. Disponível em: <<https://www.degruyter.com/view/journals/jag/2/2/article-p67.xml>>. Citado na página 55.
- KOCH, K. R. Expectation maximization algorithm and its minimal detectable outliers. *Stud. Geophys. Geod.*, v. 61, p. 1–18, 2017. Citado na página 56.
- KOK, M. V.; KAYA, E.; AKIN, S. Monte Carlo simulation of oil fields. *Energy Sources, Part B: Economics, Planning, and Policy*, Taylor Francis, v. 1, n. 2, p. 207–211, 2006. Disponível em: <<https://doi.org/10.1080/15567240500400770>>. Citado na página 54.
- LEHMANN, E. L.; ROMANO, J. P. *Testing Statistical Hypotheses*. 3. ed. [S.l.]: Springer-Verlag New York, 2005. Citado na página 40.

LEHMANN, R. Improved critical values for extreme normalized and studentized residuals in Gauss–Markov models. *J. Geod.*, v. 86, n. 12, p. 1137–1146, Dec 2012. ISSN 1432-1394. Disponível em: <<https://doi.org/10.1007/s00190-012-0569-0>>. Citado 7 vezes nas páginas 24, 28, 40, 51, 55, 59 e 69.

LEHMANN, R. 3σ -rule for outlier detection from the viewpoint of geodetic adjustment. *Journal of Surveying Engineering*, v. 139, n. 4, p. 157–165, 2013. Disponível em: <<https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29SU.1943-5428.0000112>>. Citado na página 26.

LEHMANN, R. On the formulation of the alternative hypothesis for geodetic outlier detection. *J. Geod.*, v. 87, n. 4, p. 373–386, Apr 2013. ISSN 1432-1394. Disponível em: <<https://doi.org/10.1007/s00190-012-0607-y>>. Citado 3 vezes nas páginas 24, 25 e 55.

LEHMANN, R. Observation error model selection by information criteria vs. normality testing. *Studia Geophysica et Geodaetica*, v. 59, n. 4, p. 489–504, Oct 2015. ISSN 1573-1626. Disponível em: <<https://doi.org/10.1007/s11200-015-0725-0>>. Citado 5 vezes nas páginas 26, 35, 55, 219 e 220.

LEHMANN, R.; LÖSLER, M. Multiple outlier detection: Hypothesis tests versus model selection by information criteria. *J. Surv. Eng.*, v. 142, n. 4, p. 04016017, 2016. Disponível em: <<https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29SU.1943-5428.0000189>>. Citado 2 vezes nas páginas 41 e 200.

LEHMANN, R.; LÖSLER, M. Congruence analysis of geodetic networks – hypothesis tests versus model selection by information criteria. *J. Appl. Geod.*, v. 11, n. 4, p. 271–283, 2017. Citado na página 55.

LEHMANN, R.; NEITZEL, F. Testing the compatibility of constraints for parameters of a geodetic adjustment model. *Journal of Geodesy*, v. 87, n. 6, p. 555–566, 2013. ISSN 1432-1394. Disponível em: <<https://doi.org/10.1007/s00190-013-0627-2>>. Citado na página 163.

LEHMANN, R.; SCHEFFLER, T. Monte Carlo based data snooping with application to a geodetic network. *J. Appl. Geod.*, v. 5, n. 3-4, p. 123–134, 2011. ISSN 1862-9024. <<https://doi.org/10.1515/JAG.2011.014>>. Citado na página 55.

LEHMANN, R.; VOß-BÖHME, A. On the Statistical Power of Baarda’s Outlier Test and Some Alternative. *Journal of Geodetic Science*, De Gruyter, Berlin, Boston, v. 7, n. 1, p. 68 – 78, 2017. Disponível em: <<https://www.degruyter.com/view/journals/jogs/7/1/article-p68.xml>>. Citado na página 24.

LEICK, A.; RAPOPORT, L.; TATARNIKOV, D. *GPS Satellite Surveying*. [S.l.]: John Wiley Sons, 2015. ISBN 9781119018612. Citado 2 vezes nas páginas 24 e 28.

MATSUMOTO, M.; NISHIMURA, T. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, v. 8, n. 1, p. 3–30, Jan 1998. ISSN 1049-3301. <<https://dl.acm.org/citation.cfm?id=272991>>. Citado na página 58.

MATSUOKA, M. T. *Influência de diferentes condições da ionosfera no posicionamento por ponto com GPS: avaliação na região brasileira*. Tese (Doutorado) — Programa de Pós-Graduação em Ciências Cartográficas, Universidade Estadual Paulista, Faculdade

de Ciências e Tecnologia, 2007. Disponível em: <<http://hdl.handle.net/11449/100257>>. Citado na página 24.

MATSUOKA, M. T. et al. Control points selection based on maximum external reliability for designing geodetic networks. *Applied Sciences*, v. 10, n. 2, 2020. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/10/2/687>>. Citado na página 40.

METROPOLIS, N. The beginning of the Monte Carlo method. In: . Los Alamos Science (Special Issue), 1987. v. 15, p. 125–10. Disponível em: <<https://dasher.wustl.edu/chem478/reading/lascience-14-125-87.pdf>>. Citado na página 55.

MIERLO, J. V. Statistical Analysis of Geodetic Networks Designed for the Detection of Crustal Movements. In: BORRADAILE, G. et al. (Ed.). *Progress in geodynamics*. [S.l.: s.n.], 1975. p. 52–61. Citado na página 69.

MILLER, R. G. J. *Simultaneous Statistical Inference*. 2. ed. [S.l.]: Springer-Verlag New York, 1981. Citado na página 40.

MONICO, J. *Posicionamento pelo GNSS: descrição, fundamentos e aplicações*. 2. ed. [S.l.]: Editora Unesp, 2008. Citado na página 28.

NIEMEIER, W.; TENGEN, D. Uncertainty assessment in geodetic network adjustment by combining gum and Monte-Carlo-simulations. *Journal of Applied Geodesy*, De Gruyter, Berlin, Boston, v. 11, n. 2, p. 67 – 76, 2017. Disponível em: <<https://www.degruyter.com/view/journals/jag/11/2/article-p67.xml>>. Citado na página 55.

PRÓSZYŃSKI, W. Revisiting Baarda's concept of minimal detectable bias with regard to outlier identifiability. *J. Geod.*, v. 89, n. 10, p. 993–1003, Oct 2015. ISSN 1432-1394. Disponível em: <<https://doi.org/10.1007/s00190-015-0828-y>>. Citado 2 vezes nas páginas 45 e 69.

ROBERT, C.; CASELLA, G. *Monte Carlo Statistical Methods*. 2. ed. [S.l.]: Springer-Verlag New York, 2004. ISSN 1431-875X. ISBN 978-0-387-21239-5. Citado na página 54.

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. Disponível em: <<http://dx.doi.org/10.4314/sajg.v6i3.11>>. Citado 4 vezes nas páginas 25, 55, 59 e 68.

ROFATTO, V.; MATSUOKA, M.; KLEIN, I. Design of geodetic networks based on outlier identification criteria: An example applied to the leveling network. *Bull. Geod. Sci.*, v. 24, n. 2, p. 152–170, 2018. <<http://dx.doi.org/10.1590/s1982-21702018000200011>>. Citado na página 55.

ROFATTO, V. F. et al. Monte-Carlo-based uncertainty propagation in the context of Gauss–Markov model: a case study in coordinate transformation. *Scientia Plena*, v. 15, n. 9, p. 1–17, 2019. Citado 4 vezes nas páginas 27, 35, 219 e 220.

ROFATTO, V. F. et al. A half-century of Baarda's concept of reliability: a review, new perspectives, and applications. *Survey Review*, Taylor Francis, v. 52, n. 372, p. 261–277, 2020. Disponível em: <<https://doi.org/10.1080/00396265.2018.1548118>>. Citado 5 vezes nas páginas 24, 45, 55, 59 e 66.

- ROFATTO, V. F. et al. A Monte Carlo-based Outlier Diagnosis Method for Sensitivity Analysis. *Remote Sensing*, v. 12, n. 5, 2020. ISSN 2072-4292. Disponível em: <<https://www.mdpi.com/2072-4292/12/5/860>>. Citado 3 vezes nas páginas 66, 117 e 168.
- ROM, D. M. An Improved Hochberg Procedure for Multiple Tests of Significance. *The British journal of mathematical and statistical psychology*, v. 66, n. 1, p. 189â”196, February 2013. ISSN 0007-1102. Disponível em: <<https://doi.org/10.1111/j.2044-8317.2012.02042.x>>. Citado na página 40.
- ROUSSEEUW, P. J.; LEROY, A. M. *Robust Regression and Outlier Detection*. 1. ed. [S.l.]: Wiley-Interscience, 2003. ISBN 978-0471488552. Citado na página 25.
- SANTOS, W. et al. Using the Monte Carlo technique to calculate dose conversion coefficients for medical professionals in interventional radiology. *Radiation Physics and Chemistry*, v. 95, p. 177 – 180, 2014. ISSN 0969-806X. Proceedings of the 12th International Symposium on Radiation Physics (ISRP 2012). Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0969806X13000510>>. Citado na página 54.
- SARKAR, S. K.; CHANG, C.-K. The Simes Method for Multiple Hypothesis Testing with Positively Dependent Test Statistics. *Journal of the American Statistical Association*, Taylor Francis, v. 92, n. 440, p. 1601–1608, 1997. Citado na página 40.
- SCHAFFRIN, B. Reliability measures for correlated observations. *Journal of Surveying Engineering*, v. 123, n. 3, p. 126–137, 1997. Citado na página 69.
- SIMES, R. J. An Improved Bonferroni Procedure for Multiple Tests of Significance. *Biometrika*, v. 73, n. 3, p. 751–754, 12 1986. ISSN 0006-3444. Disponível em: <<https://doi.org/10.1093/biomet/73.3.751>>. Citado na página 40.
- STIGLER, S. *Statistics on the Table: The History of Statistical Concepts and Methods*. 1st. ed. [S.l.]: Harvard University Press., 2002. Citado na página 54.
- SURACI, S. S.; OLIVEIRA, L. C. de. Chebyshev norm adjustment: what to expect? case study in a leveling network. *Revista Brasileira de Geomática*, v. 7, n. 4, p. 172–185, 2010. Citado 2 vezes nas páginas 25 e 55.
- TANIZAKI, H. *Computational Methods in Statistics and Econometrics*. 1st. ed. [S.l.]: Marcel Dekker, 2004. Citado na página 67.
- TEUNISSEN, P. An integrity and quality control procedure for use in multi sensor integration. In: . [S.l.: s.n.], 1990. Citado na página 39.
- TEUNISSEN, P. *Testing Theory: an introduction*. 2. ed. [S.l.]: Delft University Press, 2006. ISBN 9040719756. Citado 4 vezes nas páginas 26, 28, 37 e 213.
- TEUNISSEN, P. J. G. Distributional theory for the DIA method. *Journal of Geodesy*, v. 92, n. 1, p. 59–80, Jan 2018. ISSN 1432-1394. Disponível em: <<https://doi.org/10.1007/s00190-017-1045-7>>. Citado 4 vezes nas páginas 26, 52, 69 e 198.

- TEUNISSEN, P. J. G.; AMIRI-SIMKOOEI, A. R. Least-squares variance component estimation. *Journal of Geodesy*, v. 82, n. 2, p. 65–82, 2008. ISSN 1432-1394. Disponível em: <<https://doi.org/10.1007/s00190-007-0157-x>>. Citado na página 214.
- TEUNISSEN, P. J. G.; IMPARATO, D.; TIBERIUS, C. C. J. M. Does RAIM with correct exclusion produce unbiased positions? *Sensors*, v. 17, n. 7, 2017. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/17/7/1508>>. Citado na página 30.
- TIENSTRA, J. M. *Theory of the adjustment of normally distributed observations*. Amsterdam : Argus, 1956. Disponível em: <<http://lib.ugent.be/catalog/rug01:000474849>>. Citado na página 50.
- VANÍČEK, P.; CRAYMER, M. R.; KRAKIWSKY, E. J. Robustness analysis of geodetic horizontal networks. *J. Geod.*, v. 75, n. 4, p. 199–209, Jul 2001. ISSN 1432-1394. Disponível em: <<https://doi.org/10.1007/s001900100162>>. Citado na página 39.
- VELSINK, H. On the deformation analysis of point fields. *J. Geod.*, v. 89, n. 11, p. 1071–1087, Nov 2015. ISSN 1432-1394. Disponível em: <<https://doi.org/10.1007/s00190-015-0835-z>>. Citado na página 163.
- WIEGAND, R. E. Introductory econometrics: Using Monte Carlo simulation with microsoft excel. *Journal of the American Statistical Association*, Taylor Francis, v. 101, n. 476, p. 1730–1731, 2006. Disponível em: <<https://doi.org/10.1198/jasa.2006.s154>>. Citado na página 54.
- WILCOX, R. *Introduction to Robust Estimation and Hypothesis Testing*. 3. ed. [S.l.]: Academic Press, 2013. ISBN 978-0-12-386983-8. Citado na página 25.
- WRIGHT, S. P. Adjusted p-values for simultaneous inference. *Biometrics*, [Wiley, International Biometric Society], v. 48, n. 4, p. 1005–1013, 1992. ISSN 0006341X, 15410420. Disponível em: <<http://www.jstor.org/stable/2532694>>. Citado na página 40.
- Wyszkowska, P. Propagation of uncertainty by Monte Carlo simulations in case of basic geodetic computations. *Geodesy and Cartography*, v. 66, n. 2, p. 333–346, dez. 2017. Citado na página 55.
- YANG, L. et al. Outlier separability analysis with a multiple alternative hypotheses test. *J. Geod.*, v. 87, n. 6, p. 591–604, Jun 2013. ISSN 1432-1394. Disponível em: <<https://doi.org/10.1007/s00190-013-0629-0>>. Citado 5 vezes nas páginas 25, 28, 45, 69 e 197.
- YANG, Y. Robust estimation of geodetic datum transformation. *J. Geod.*, v. 73, n. 5, p. 268–274, Jun 1999. ISSN 1432-1394. Disponível em: <<https://doi.org/10.1007/s001900050243>>. Citado na página 25.
- ZAMINPARDAZ, S.; TEUNISSEN, P. DIA-datasnooping and identifiability. *J. Geod.*, v. 93, n. 1, p. 85–101, Jan 2019. ISSN 1432-1394. Disponível em: <<https://doi.org/10.1007/s00190-018-1141-3>>. Citado 3 vezes nas páginas 69, 198 e 200.
- ZAMINPARDAZ, S.; TEUNISSEN, P. How Abnormal Are the PDFs of the DIA Method: A Quality Description in the Context of GNSS. In: _____. *International Association of Geodesy Symposia*. [S.l.]: Springer, Berlin, Heidelberg, 2019. p. 1–19. Citado 2 vezes nas páginas 53 e 200.

Apêndices

APÊNDICE A – Teoria da Estimação

Todo processo de medição, ou coleta de dados, tem imperfeições que dão origem a incertezas. Mesmo que se utilizem os mais sofisticados equipamentos sob condições controladas de trabalho, os resultados que se obtêm virão afetados por diversas fontes de incerteza. Na prática, existem inúmeras fontes possíveis de incerteza em uma medição. Aqui listamos algumas:

1. Definição incompleta da grandeza;
2. Falhas na definição da grandeza;
3. Conhecimento inadequado dos efeitos das condições ambientais sobre a medição ou medição imperfeita das condições ambientais;
4. Erro de tendência pessoal na leitura de instrumentos analógicos;
5. Resolução finita do instrumento;
6. Valores inexatos de constantes e de outros parâmetros obtidos de fontes externas e usados no algoritmo para obtenção de dados;
7. Aproximações e suposições incorporadas ao método e procedimento de medição;
8. Variações nas observações repetidas da grandeza sob condições aparentemente idênticas.

Essas fontes não são necessariamente independentes. Por exemplo, as fontes (3) e (7) podem contribuir para a fonte (8). De modo geral, a incerteza do resultado de uma medição reflete a falta de conhecimento associado ao valor da grandeza a ser medida. Quanto maior for a incerteza, tanto menor será a confiabilidade do resultado da medição.

Como as fontes de incerteza são inevitáveis, associamos uma variável aleatória (ou vetor aleatório) para descrever certos fenômenos físicos que se deseja estudar. Uma variável aleatória é o resultado de um evento sujeito a incertezas, ou seja, de um processo experimental de medição. Uma variável aleatória pode ser de natureza discreta ou contínua. Aqui, todas as variáveis aleatórias são tratadas como de natureza contínua. Um vetor aleatório \mathbf{y} é um vetor cujos componentes são variáveis aleatórias.

A aleatoriedade de \mathbf{y} é expressa por sua função de densidade de probabilidade (f.d.p.). A distribuição normal multivariada, por exemplo, descreve uma população de eventos. Uma população é a totalidade de todos os eventos. Observa-se que é impraticável

na grande maioria dos casos, estudar-se a população. Portanto, nosso conhecimento da f.d.p. de \mathbf{y} é incompleto. A f.d.p. geralmente pode ser indexada com um ou mais parâmetros desconhecidos. A f.d.p. de um vetor aleatório \mathbf{y} é denotado por $f_{\mathbf{y}}(\mathbf{y}|\mathbf{x})$, na qual \mathbf{x} é um vetor de parâmetros desconhecidos a ser estimado. Importante notar que $\mathbf{y} \in \mathbb{R}^n$, enquanto que $\mathbf{x} \in \mathbb{R}^u$. Na impossibilidade de termos todos os valores possíveis de \mathbf{y} , uma pequena parcela de \mathbf{y} , denominada amostra, é selecionada. Essa amostra é denotada por $\bar{\mathbf{y}}$. O vetor de observações $\bar{\mathbf{y}}$ como uma realização de \mathbf{y} com f.d.p. $f_{\mathbf{y}}(\mathbf{y}|\mathbf{x})$ contém informações que podem ser usadas para estimar \mathbf{x} .

O problema é, portanto, encontrar uma função $\mathcal{F} : \mathbb{R}^n \mapsto \mathbb{R}^u$, de modo que $\hat{\mathbf{x}} = \mathcal{F}(\bar{\mathbf{y}})$ possa ser tomada como nossa estimativa de \mathbf{x} . Na verdade, se aplicarmos \mathcal{F} em \mathbf{y} , nós obtemos o vetor aleatório $\hat{\mathbf{x}} = \mathcal{F}(\mathbf{y})$. O vetor aleatório $\hat{\mathbf{x}}$ é chamado estimador de \mathbf{x} e $\hat{\mathbf{x}}$ é chamado estimativa de \mathbf{x} . O estimador é um vetor aleatório que possui sua própria f.d.p. (distribuição), enquanto que a estimativa é um valor realizado do estimador e, portanto, um vetor determinístico. A diferença $\hat{\boldsymbol{\epsilon}} = \hat{\mathbf{x}} - \mathbf{x}$ é chamada de erro do estimador. Como $\hat{\mathbf{x}}$ depende da escolha da função \mathcal{F} , o erro do estimador também dependerá de \mathcal{F} . Listamos três propriedades desejáveis para $\hat{\boldsymbol{\epsilon}}$, as quais expressam, em certo sentido, a proximidade de $\hat{\mathbf{x}}$ à \mathbf{x} . Tais propriedades podem ser usadas como critério para encontrar uma função \mathcal{F} “ótima”.

A.1 Propriedades ótimas dos estimadores

Imparcialidade(*Unbiasedness*): o estimador $\hat{\mathbf{x}}$ é dito ser um estimador imparcial, ou não tendencioso, de \mathbf{x} se e somente se a esperança matemática do erro do estimador é zero, i.e. $\mathbb{E}\{\hat{\boldsymbol{\epsilon}}\} = 0$. Um estimador é, portanto, imparcial se a média de sua distribuição for igual a \mathbf{x} , ou seja:

$$\mathbb{E}\{\hat{\mathbf{x}}\} = \mathbf{x} \text{ para todo } \mathbf{x} \quad (\text{A.1})$$

Isso implica que a média de realizações repetidas de $\hat{\boldsymbol{\epsilon}}$ tenderão a zero. Um estimador que não é imparcial é dito ser tendencioso, e a diferença $\mathbb{E}\{\hat{\boldsymbol{\epsilon}}\} = \mathbb{E}\{\hat{\mathbf{x}}\} - \mathbf{x}$ é chamada de bias (viés) do estimador. O tamanho do bias é, portanto, uma medida da proximidade de $\hat{\mathbf{x}}$ à \mathbf{x} . O erro médio $\mathbb{E}\{\hat{\boldsymbol{\epsilon}}\}$ é uma medida de proximidade que faz uso do primeiro momento da distribuição de $\hat{\mathbf{x}}$.

Variância Mínima(*best*): uma segunda medida de proximidade do estimador a \mathbf{x} é o *erro quadrático médio* (mean squared error, em inglês – *MSE*), que é definido como:

$$MSE = \mathbb{E}\{||\hat{\mathbf{x}} - \mathbf{x}||^2\} \rightarrow \min, \quad (\text{A.2})$$

em que $\|\cdot\|$ é a norma de um vetor. Se comparássemos diferentes estimadores, observando suas respectivas MSEs, preferiríamos aquele com o menor MSE. Essa é uma medida de proximidade que faz uso também do segundo momento da distribuição de $\hat{\mathbf{x}}$. Portanto, o melhor estimador (*best*) é, na ausência de tendência (viés ou bias), aquele de variação mínima.

Máxima Verossimilhança (*Maximum likelihood*): em vez de confiar nos dois primeiros momentos de uma distribuição, pode-se definir a proximidade em termos da própria distribuição. Como terceira medida de proximidade, consideramos, portanto, a probabilidade do estimador $\hat{\mathbf{x}}$ residir em uma pequena região centralizada em \mathbf{x} . Se considerarmos que esta região é uma hipersfera com um dado raio R , a medida será dada como:

$$\mathcal{P}(\|\hat{\mathbf{x}} - \mathbf{x}\|^2 \leq R^2) \rightarrow \max \quad (\text{A.3})$$

Se comparássemos diferentes estimadores, observando seus respectivos valores de probabilidade em A.3, preferiríamos aquele com a maior probabilidade. Infelizmente, raramente é possível derivar um estimador com a maior probabilidade para todos os \mathbf{x} .

Existem diferentes métodos de estimativa que discutimos neste capítulo. O método mais simples que se pode aplicar precisa de informações sobre o primeiro momento da distribuição. Como a f.d.p. depende do parâmetro desconhecido \mathbf{x} , a média também, geralmente, depende de \mathbf{x} . No entanto, assumiremos que sabemos como a média depende de \mathbf{x} . Em outras palavras, esta relação é dada por meio de um mapeamento: $\mathcal{A} : \mathbb{R}^n \mapsto \mathbb{R}^u$. No caso linear (ou linearizado), \mathbf{A} é uma matriz, conhecida como matriz *design*. Por esse motivo, o problema é chamado de *A-model*.

Redundância: frequentemente, são feitas medições redundantes para diminuir a incerteza dos resultados obtidos e verificar a presença de *outliers*, i.e. $n > u$, em que n é a dimensão do espaço das observações, ou simplesmente número de observações, e u refere-se à dimensão do espaço dos parâmetros, i.e. número de parâmetros desconhecidos.

Devido à incerteza intrínseca nas observações, medições redundantes geralmente levam a um sistema inconsistente de equações. Para $n > u$ tal sistema linear inconsistente de equações tem a forma $\mathbf{y} \approx \mathbf{Ax}$. Esse sistema de equações para o qual $n > u = \text{car}(\mathbf{A})$ também é referido como um sistema sobre-determinado. O número $r = n - \text{car}(\mathbf{A}) = n - u$ é chamado de redundância do sistema (ou redundância do modelo funcional).

Princípio dos Mínimos Quadrados: é sabido que um sistema inconsistente não tem solução \mathbf{x} que possa reproduzir \mathbf{y} . Como primeiro passo, é possível tornar o sistema consistente introduzindo um vetor de incertezas proveniente dos efeitos aleatórios e da correção imperfeita para efeitos sistemáticos \mathbf{e} : $\mathbf{y} = \mathbf{Ax} + \mathbf{e}$. Mas agora temos $n + u$ incógnitas em n equações e, portanto, um número infinito de soluções possíveis para \mathbf{x} e \mathbf{e}

(sistema sub-determinado). Parece razoável selecionar deste conjunto infinito de soluções possíveis, a solução que, em certo sentido, fornece o menor valor para \mathbf{e} . Isso implica escolher a solução para \mathbf{x} de modo que \mathbf{Ax} seja, em algum sentido, o mais próximo possível do vetor das observações \mathbf{y} . O princípio dos mínimos quadrados é minimizar a norma (ponderada) do vetor dos resíduos \mathbf{e} , ou seja:

$$\|\mathbf{e}\|_W^2 = \mathbf{e}^T \mathbf{W} \mathbf{e} = (\mathbf{y} - \mathbf{Ax})^T \mathbf{W} (\mathbf{y} - \mathbf{Ax}) \rightarrow \min \quad (\text{A.4})$$

em que \mathbf{W} é a matriz peso. Qualquer matriz simétrica positiva-definida é considerada estar na classe de matriz de peso admissível \mathbf{W} .

Modelo linear: a partir de agora, nós referiremos ao sistema linear de equações $\mathbf{y} = \mathbf{Ax} + \mathbf{e}$ como o modelo linear de equações de observação, o qual é dado por:

$$\mathbb{E}\{\mathbf{y}\} = \mathbf{Ax}, \quad \mathbb{D}\{\mathbf{y}\} = \mathbf{Q}_e, \quad \text{ou } \mathbf{W} \quad (\text{A.5})$$

em que $\mathbf{y} \in \mathbb{R}^{n \times 1}$ é o vetor das observações, $\mathbf{A} \in \mathbb{R}^{n \times u}$ é a matriz *design*, $\mathbf{x} \in \mathbb{R}^{u \times 1}$ é o vetor dos parâmetros incógnitos, $\mathbf{W} \in \mathbb{R}^{n \times n}$ é a matriz peso e $\mathbf{Q}_e \in \mathbb{R}^{n \times n}$ é a matriz covariância das observáveis. Assume-se que a matriz design \mathbf{A} tenha posto completo (*full column rank*), i.e. $\text{car}(\mathbf{A}) = u$, e portanto, temos $n > u$. As matrizes \mathbf{W} e \mathbf{Q}_e são simétricas positivas-definidas. Novamente, $\mathbb{E}\{\cdot\}$ é o operador de esperança matemática e $\mathbb{D}\{\cdot\}$ é o operador de dispersão.

Métodos de Estimação: três métodos diferentes de estimação serão tratados neste capítulo. São eles: estimação por mínimos quadrados ponderados (WLSE – Weighted Least-Squares), melhor estimador imparcial linear (BLUE – Best Linear Unbiased Estimator) e estimador de máxima verossimilhança (MLE – Maximum Likelihood Estimation). Os métodos diferem não apenas nos princípios de estimação envolvidos, mas também nas informações necessárias sobre a f.d.p. $\mathbf{f}_y(\mathbf{y}|\mathbf{x})$.

O WLSE é aplicado quando temos apenas informações sobre o primeiro momento da distribuição. Importante mencionar que não é necessário fazer qualquer inferência sobre os dados observados para aplicar o WLSE. O BLUE é um método que pode ser aplicado quando tivermos informações sobre os dois primeiros momentos da distribuição. O MLE é usado se conhecermos a estrutura completa de $\mathbf{f}_y(\mathbf{y}|\mathbf{x})$. Um exemplo importante para o qual a estrutura completa da f.d.p. é conhecida é a distribuição normal multivariada, ou seja, $\mathbf{y} \sim \mathcal{N}(\mathbf{Ax}, \mathbf{Q}_e)$.

A.2 Método dos mínimos quadrados ponderados

Definição: seja $\mathbb{E}\{\mathbf{y}\} = \mathbf{Ax}$ um modelo linear possivelmente inconsistente de equações de observação, com $\mathbf{A} \in \mathbb{R}^{n \times u}$, tendo posto completo, ou seja, $\text{car}(\mathbf{A}) = u$.

Suponha também que $\mathbf{W} \in \mathbb{R}^{n \times n}$ seja uma matriz simétrica positiva-definida, ou seja, $\mathbf{z}^T \mathbf{W} \mathbf{z} > 0$, para todo vetor \mathbf{z} de dimensão $n \times 1$. Então, a solução dos mínimos quadrados ponderados (WLSE) é dada por:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^u} (\mathbf{y} - \mathbf{A}\mathbf{x})^T \mathbf{W} (\mathbf{y} - \mathbf{A}\mathbf{x}) \quad (\text{A.6})$$

A diferença $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}$ é chamada de vetor dos resíduos dos mínimos quadrados (ponderados). Sua norma ao quadrado (ponderada) $\|\hat{\mathbf{e}}\|_W^2 = \hat{\mathbf{e}}^T \mathbf{W} \hat{\mathbf{e}}$ é uma medida escalar para a inconsistência do modelo linear. Ela é utilizada como estatística de teste na validação global do modelo (*overall model test*).

Estimador: como a média de \mathbf{y} depende do desconhecido \mathbf{x} , a f.d.p. de \mathbf{y} também depende do desconhecido \mathbf{x} . O problema de determinar um valor para \mathbf{x} pode agora ser visto como um problema de estimação, ou seja, como o problema de encontrar uma função \mathcal{F} de modo que $\hat{\mathbf{x}} = \mathcal{F}(\bar{\mathbf{y}})$ atue como estimativa de \mathbf{x} e $\hat{\mathbf{x}} = \mathcal{F}(\mathbf{y})$ como estimador de \mathbf{x} . O estimador dos mínimos quadrados ponderados (WLSE) é dado por:

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{y} \quad (\text{A.7})$$

$\hat{\mathbf{x}}$ é um estimador linear, pois todos os elementos de $\hat{\mathbf{x}}$ são combinações lineares dos elementos de $\hat{\mathbf{y}}$. O estimador dos mínimos quadrados $\hat{\mathbf{y}} = \mathbf{A}\hat{\mathbf{x}}$ das observáveis e $\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}}$ dos resíduos se originam a partir da relação $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ como:

$$\begin{cases} \hat{\mathbf{y}} = P_A \mathbf{y} \\ \hat{\mathbf{e}} = P_A^\perp \mathbf{y} \end{cases} \quad (\text{A.8})$$

com $P_A = \mathbf{A}(\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W}$ e $P_A^\perp = I - P_A$ sendo dois projetores ortogonais. I é uma matriz identidade, i.e. $I \in \mathcal{R}^{n \times n}$. O projetor P_A projeta sobre o espaço coluna da matriz \mathbf{A} (denotado por $\mathcal{R}(\mathbf{A})$) e na direção de seu complemento ortogonal (i.e. $\mathcal{R}(\mathbf{A})^\perp$), enquanto que P_A^\perp projeta sobre $\mathcal{R}(\mathbf{A})^\perp$ e na direção de $\mathcal{R}(\mathbf{A})$. Para algumas propriedades úteis desses dois projetores, consulte (TEUNISSEN, 2006). $\mathcal{R}(\cdot)$ denota o espaço coluna de uma matriz. Importante destacar que projetor P_A^\perp é conhecido como matriz redundância, denotado aqui por \mathbf{R} , i.e. $\mathbf{R} = I - \mathbf{A}(\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W}$. A diagonal da matriz \mathbf{R} é composto por elementos conhecidos como número de redundância local, denotado por r_i . Par

Imparcialidade: para obter algumas compreensão sobre o desempenho de um estimador, precisamos saber como o estimador se relaciona com seu valor-alvo. Com base no pressuposto $\mathbb{E}\{\mathbf{e}\} = 0$, as esperanças de $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ e $\hat{\mathbf{e}}$ são dadas, respectivamente, por:

$$\mathbb{E}\{\hat{\mathbf{x}}\} = \mathbf{x}, \quad \mathbb{E}\{\hat{\mathbf{y}}\} = \mathbb{E}\{\mathbf{y}\} = \mathbf{A}\mathbf{x}, \quad \mathbb{E}\{\hat{\mathbf{e}}\} = \mathbb{E}\{\mathbf{e}\} = 0 \quad (\text{A.9})$$

Se $\mathbb{E}\{e$, então o WLSE é um estimador imparcial linear. A imparcialidade é claramente uma propriedade desejável. Isso implica que, em média, os resultados do estimador estarão no alvo. Observe que a imparcialidade do WLSE é independente da escolha feita para a matriz dos pesos \mathbf{W} .

Matriz Covariância: para obter as matrizes de covariância de $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ e $\hat{\mathbf{e}}$, precisamos da matriz de covariância das observáveis \mathbf{Q}_e . As matrizes de covariância de $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ e $\hat{\mathbf{e}}$ serão indicadas respectivamente por $\mathbf{Q}_{\hat{\mathbf{x}}}$, $\mathbf{Q}_{\hat{\mathbf{y}}}$ e $\mathbf{Q}_{\hat{\mathbf{e}}}$. Aplicando a *lei de propagação dos erros* nas Equações A.7 e A.8 produz:

$$\begin{cases} \mathbf{Q}_{\hat{\mathbf{x}}} = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{Q}_e \mathbf{W} \mathbf{A} (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \\ \mathbf{Q}_{\hat{\mathbf{y}}} = P_A \mathbf{Q}_e P_A^\perp \\ \mathbf{Q}_{\hat{\mathbf{e}}} = P_A^\perp \mathbf{Q}_e P_A^{\perp T} \end{cases} \quad (\text{A.10})$$

As matrizes covariâncias dependem da escolha feita para a matriz \mathbf{W} .

Erro Quadrático Médio: como visto anteriormente, o tamanho do erro do estimador, o qual foi chamado de *mean squares error* (MSE), pode ser quantificado por meio da soma dos quadrados seus elementos, i.e. $\mathbb{E}\{\hat{\mathbf{e}}^T \hat{\mathbf{e}}\} = \mathbb{E}\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\}$, o qual foi chamado de *mean squares error* (MSE). Pode-se facilmente mostrar que o MSE é decomposto como $\mathbb{E}\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\} = \mathbb{E}\{\|\hat{\mathbf{x}} - \mathbb{E}\{\hat{\mathbf{x}}\}\|^2\} + \mathbb{E}\{\|\mathbf{x} - \mathbb{E}\{\hat{\mathbf{x}}\}\|^2\}$. O primeiro termo no lado direito é o traço da matriz covariância do estimador e o segundo termo é a norma quadrática da tendência do estimador (bias ou viés). Como o WLSE é imparcial (não tendencioso), o segundo termo desaparece e, portanto, o MSE do WLSE é dado por:

$$\mathbb{E}\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\} = \text{tr}(\mathbf{Q}_{\hat{\mathbf{x}}}) \quad (\text{A.11})$$

em que $\text{tr}(\cdot)$ é o traço de uma matriz.

Forma Quadrática dos resíduos: nos mínimos quadrados ponderados (WLSE), um critério importante que mostra a inconsistência do modelo linear de equações de observação é a forma quadrática (norma quadrática) dos resíduos, a qual foi dada em 2.2 e reescrita como:

$$\|\hat{\mathbf{e}}\|_W^2 = \hat{\mathbf{e}}^T \mathbf{W} \hat{\mathbf{e}} = \mathbf{y}^T \mathbf{W} \mathbf{y} - \mathbf{y}^T \mathbf{W} \mathbf{A} (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{y} \quad (\text{A.12})$$

Os resultados nas equações (A.10), (A.11) e (A.12) são todos independentes da distribuição de \mathbf{y} . A propriedade de imparcialidade em (A.9) é ainda independente da escolha feita para a matriz dos pesos \mathbf{W} , enquanto as matrizes de covariância (A.10), o erro quadrático médio (A.11) e a forma quadrática (A.12) dependem de \mathbf{W} . Pode-se, portanto, pensar em selecionar uma “melhor” matriz dos pesos, ou seja, uma matriz dos pesos que minimiza o MSE (TEUNISSEN; AMIRI-SIMKOOEI, 2008).

Normalidade: os vetores aleatórios $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ e $\hat{\mathbf{e}}$ são todos funções lineares de $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ e \mathbf{y} . Isso implica que eles tenham uma distribuição gaussiana (normal) sempre que \mathbf{y} tiver uma distribuição normal. Em outras palavras, se \mathbf{y} tem distribuição normal $\mathbf{y} \sim \mathcal{N}(\mathbf{Ax}, \mathbf{Q}_e)$, então:

$$\hat{\mathbf{x}} \sim \mathcal{N}(\mathbf{x}, \mathbf{Q}_{\hat{\mathbf{x}}}), \quad \hat{\mathbf{y}} \sim \mathcal{N}(\mathbf{Ax}, \mathbf{Q}_{\hat{\mathbf{y}}}), \quad \hat{\mathbf{e}} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{\hat{\mathbf{e}}}) \quad (\text{A.13})$$

Observe que, como a matriz de covariância $\mathbf{Q}_{\hat{\mathbf{e}}}$ é singular, o vetor dos resíduos dos mínimos quadrados $\hat{\mathbf{e}}$ tem uma distribuição normal singular. Observe também que as duas primeiras distribuições diferem da terceira no sentido de que suas médias são desconhecidas. Portanto, se \mathbf{Q}_e é dado, apenas a forma de suas f.d.p. é conhecida, mas não a localização. A f.d.p. de $\hat{\mathbf{e}}$, entretanto, é completamente conhecida, uma vez que o modelo de incerteza é fornecido, \mathbf{Q}_e . Esta última propriedade teve um papel fundamental para o método desenvolvido nessa Tese.

A.3 Melhor estimador linear imparcial

Mínima Variância: a abordagem dos mínimos quadrados (ponderados) foi introduzida como uma técnica atraente para resolver um sistema inconsistente de equações. O método em si é determinístico, uma vez que não são utilizados conceitos da teoria da probabilidade na formulação do problema de minimização. Para selecionar um estimador ideal da classe de estimadores imparciais lineares (LUE – Linear Unbiased Estimator), precisamos definir o critério de otimização. Como critério, escolhemos a minimização do erro quadrático médio (MSE). O estimador que possui o menor MSE de todas as LUEs é chamado de melhor estimador linear imparcial (BLUE). Esse problema de minimização resulta na menor variância possível para estimadores, ou seja, $\mathbb{E}\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\} = \text{tr}(\mathbf{Q}_{\hat{\mathbf{x}}}) \rightarrow \min$.

Estimador: se a matriz covariância \mathbf{Q}_e das observáveis é conhecida, pode-se usar o (BLUE) tomando a matriz dos pesos \mathbf{W} como a inversa da \mathbf{Q}_e , ou seja, $\mathbf{W} = \mathbf{Q}_e^{-1}$. Neste caso, temos os seguintes resultados:

$$\begin{cases} \hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{Q}_e^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Q}_e^{-1} \mathbf{y} \\ \hat{\mathbf{y}} = P_A \mathbf{y} \\ \hat{\mathbf{e}} = P_A^\perp \mathbf{y} \end{cases} \quad (\text{A.14})$$

em que tanto $P_A = \mathbf{A}(\mathbf{A}^T \mathbf{Q}_e^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Q}_e^{-1}$ quanto $P_A^\perp = I - P_A$ são projetores ortogonais. Substituindo $\mathbf{W} = \mathbf{Q}_e^{-1}$ na Equação A.10, tem-se:

$$\begin{cases} \mathbf{Q}_{\hat{\mathbf{x}}} = (\mathbf{A}^T \mathbf{Q}_e^{-1} \mathbf{A})^{-1} \\ \mathbf{Q}_{\hat{\mathbf{y}}} = P_A \mathbf{Q}_e \\ \mathbf{Q}_{\hat{\mathbf{e}}} = P_A^\perp \mathbf{Q}_e \end{cases} \quad (\text{A.15})$$

Pode-se demonstrar que, de todos os estimadores lineares imparciais, o estimador BLUE possui variância mínima. Portanto, é um estimador linear imparcial de variância mínima. O BLUE é, às vezes, também chamado de estimador *probabilístico* de mínimos quadrados. A propriedade de mínima variância também é independente da distribuição de \mathbf{y} (assim, como a propriedade imparcialidade). Na literatura, a escolha de $\mathbf{W} = \mathbf{Q}_e^{-1}$ é frequentemente feita, levando ao BLUE.

Forma quadrática dos resíduos: a partir dos estimadores BLUE, o critério de inconsistência do modelo linear de equações de observação, expresso pela forma quadrática dos resíduos, é dado como:

$$\|\hat{\mathbf{e}}\|_{\mathbf{Q}_e^{-1}}^2 = \hat{\mathbf{e}}^T \mathbf{Q}_e^{-1} \hat{\mathbf{e}} = \mathbf{y}^T \mathbf{Q}_e^{-1} \mathbf{y} - \mathbf{y}^T \mathbf{Q}_e^{-1} \mathbf{A} (\mathbf{A}^T \mathbf{Q}_e^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Q}_e^{-1} \mathbf{y} \quad (\text{A.16})$$

A norma quadrada dos resíduos desempenhará um papel importante na seção de testes de hipóteses para detecção e validação do modelo.

Nos mínimos quadrados ponderados, a matriz dos pesos \mathbf{W} desempenha o papel de um tensor métrico em um espaço vetorial. Os estimadores do BLUE consideram a matriz dos peso como o inverso da matriz de covariância. Portanto, a matriz de covariância dos observáveis está intimamente relacionada ao tensor métrico. Temos, portanto, algumas interpretações probabilísticas em nosso espaço vetorial. Por exemplo, se as covariâncias entre os observáveis são zero, isso significa que os vetores de base padrão (base canônica) são ortogonais; observáveis não correlacionados significa, para vetores de base, não ter projeção um sobre o outro. Se, além disso, as variâncias são iguais, isso significa que os vetores de base são ortonormais. Portanto, se considerarmos a matriz dos pesos como a matriz inversa da covariância, a definição de distância mínima (norma mínima) no espaço vetorial obtido dos mínimos quadrados ponderados coincidirá com a definição de variância mínima no modelo estocástico (espaço) obtido do BLUE.

A.4 Estimativa de máxima verossimilhança

O WLSE e o BLUE não são baseados apenas em princípios diferentes, mas também diferem no tipo de informação exigida na f.d.p. de \mathbf{y} . Para o WLSE, precisamos apenas de informações sobre o primeiro momento do f.d.p., a média de \mathbf{y} . Para o BLUE, precisamos, além do primeiro momento, também do segundo momento da f.d.p., a matriz de covariância de \mathbf{y} . Para modelos lineares, os dois princípios, WLSE e BLUE, dão resultados idênticos quando a matriz dos pesos é tomada igual ao inverso da matriz de covariância das observáveis. Nesta seção, apresentamos o método de estimação de máxima verossimilhança (MLE), que requer completo conhecimento da f.d.p.

Princípio: o método de máxima verossimilhança (MLE – Maximum Likelihood Estimation) é conceitualmente um dos mais simples métodos de estimação. Ela só é aplicada, no entanto, quando a estrutura geral da f.d.p. é conhecida.

Suponhamos, portanto, que a f.d.p. de $\mathbf{y} \in \mathbb{R}^{n \times 1}$, i.e. $\mathbf{f}_y(\mathbf{y}|\mathbf{x})$, seja conhecida, embora com alguns u parâmetros desconhecidos. Como a f.d.p. será alterada, quando se varia \mathbf{x} , na verdade, temos uma grande família de f.d.p. em que cada membro da família é determinado pelo valor tomado por \mathbf{x} . Como \mathbf{x} é desconhecido, não se sabe à qual f.d.p. um valor observado de \mathbf{y} , y_0 , pertence. A ideia agora é selecionar, a partir das famílias de f.d.p., aquela que melhor descreve os dados observados. Para esse fim, considera-se $f_y(y_0|\mathbf{x})$ como função de \mathbf{x} . Essa função é referida como a função de probabilidade de y_0 que produz, conforme \mathbf{x} varia, as densidades de probabilidade de todas as f.d.p. para o mesmo valor de amostra y_0 . Se \mathbf{x} for o valor correto, então a probabilidade de \mathbf{y} ser um elemento de uma região infinitesimal centrada em y_0 é dada como $f_y(y_0|\mathbf{x})\partial y$. Uma escolha razoável para \mathbf{x} , dado o valor observado y_0 , é, portanto, o valor que corresponde à maior probabilidade, $\max_{\mathbf{x}} f_y(y_0|\mathbf{x})\partial y$ e, portanto, com o maior valor da função de probabilidade. O estimador de MLE de \mathbf{x} é, portanto, definido a seguir:

Definição: seja a f.d.p. do vetor das observáveis $\mathbf{y} \in \mathbb{R}^{n \times 1}$ parametrizada como $\mathbf{f}_y(\mathbf{y}|\mathbf{x})$, com $\mathbf{x} \in \mathbb{R}^{u \times 1}$ parâmetros desconhecidos. Então, o MLE de \mathbf{x} é dado como:

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x} \in \mathbb{R}^u} \mathbf{f}_y(\mathbf{y}|\mathbf{x}) \quad (\text{A.17})$$

O cálculo da solução de máxima verossimilhança em (A.17) nem sempre pode ser uma tarefa fácil. Se a função de verossimilhança é suficientemente suave, as duas condições necessárias e suficientes para que $\hat{\mathbf{x}}$ seja um maximizador (local ou global) são:

$$\partial_{\mathbf{x}} \mathbf{f}_y(\mathbf{y}|\hat{\mathbf{x}}) = \mathbf{0} \quad e \quad \partial_{\mathbf{x}\mathbf{x}^T}^2 \mathbf{f}_y(\mathbf{y}|\hat{\mathbf{x}}) < \mathbf{0} \quad (\text{A.18})$$

com $\partial_{\mathbf{x}} \mathbf{f}_y(\mathbf{y}|\hat{\mathbf{x}})$ e $\partial_{\mathbf{x}\mathbf{x}^T}^2 \mathbf{f}_y(\mathbf{y}|\hat{\mathbf{x}}) < \mathbf{0}$ sendo as derivadas de primeira e segunda ordem em relação à \mathbf{x} , respectivamente. Em outras palavras, o gradiente deve ser zero, e a matriz Hessiana (matriz simétrica das derivadas parciais de segunda ordem) deve ser negativa-definida.

No caso dos dados observados serem normalmente distribuídos, os estimadores MLE e BLUE serão idênticos. Por exemplo, considere $\mathbf{y} \sim \mathcal{N}(\mathbf{A}\mathbf{x}, \mathbf{Q}_e)$, com \mathbf{x} o vetor dos parâmetros desconhecidos. Neste caso tem-se:

$$\mathbf{f}_y(\mathbf{y}|\mathbf{x}) = (\det(2\pi\mathbf{Q}_e))^{-1/2} \exp\left(-\frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{Q}_e}^2\right), \quad (\text{A.19})$$

da qual segue que:

$$\arg \max_{\mathbf{x} \in \mathbb{R}^u} \mathbf{f}_y(\mathbf{y}|\mathbf{x}) = \arg \min_{\mathbf{x} \in \mathbb{R}^u} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{Q}_e}^2 \quad (\text{A.20})$$

Portanto, $\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{Q}_e^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Q}_e^{-1} \mathbf{y}$. Os estimadores $\hat{\mathbf{y}}$ e $\hat{\mathbf{e}}$, bem como suas respectivas matrizes de covariâncias também são coincidentes com aquelas dos estimadores BLUE.

APÊNDICE B – Testes de hipóteses em modelos lineares

Teste mais poderoso (*Most powerful test*): o teste da razão de verossimilhança simples (SLR) é derivado do princípio do teste de Neyman-Pearson. De acordo com este princípio, um teste mais poderoso (*most powerful test*) é aquele que, dentre todos os testes ou regiões críticas com o mesmo tamanho de probabilidade do Erro Tipo I, α , a probabilidade do Erro Tipo II, β , é a menor possível.

Hipóteses simples: Consideramos duas hipóteses simples. Usando o princípio de Neyman-Pearson, testaremos a hipótese nula contra a alternativa. Quando o vetor das observações \mathbf{y} possui f.d.p. conhecida $f_{\mathbf{y}}(\mathbf{y}|\mathbf{x})$, podemos definir duas hipóteses simples como: $\mathcal{H}_0 : \mathbf{x} = \mathbf{x}_0$ versus $\mathcal{H}_A : \mathbf{x} = \mathbf{x}_A$. Ambas as hipóteses pertencem a um único ponto distinto no espaço dos parâmetros. O objetivo é decidir, com base nas observações \mathbf{y} (amostra) das observáveis \mathbf{y} (população), de qual das duas distribuições as observações se originaram, de $f_{\mathbf{y}}(\mathbf{y}|\mathbf{x}_0)$ ou $f_{\mathbf{y}}(\mathbf{y}|\mathbf{x}_A)$. O teste da razão de verossimilhança simples (*simple likelihood ratio* – SLR) como regra de decisão é:

$$\text{rejeitar } \mathcal{H}_0 \text{ se } \frac{f_{\mathbf{y}}(\mathbf{y}|\mathbf{x}_0)}{f_{\mathbf{y}}(\mathbf{y}|\mathbf{x}_A)} < c, \text{ aceitar caso contrário} \quad (\text{B.1})$$

em que c é uma constante positiva (limiar, *threshold*). Pode-se provar que o SLR é o mais poderoso.

Na prática, geralmente, lidamos com hipóteses compostas. Lidamos com o problema geral de testar uma hipótese composta contra outra hipótese composta. O teste da razão de verossimilhança generalizado (*generalized likelihood ratio* – GLR) é, portanto, definido. O fato de uma hipótese composta representar mais do que apenas um ponto distinto no espaço dos parâmetros complica a noção do poder de um teste. Podemos, portanto, abordar a propriedade uniformemente mais poderosa (*uniformly most powerful* – UMP) de um teste. No caso de testar uma hipótese simples contra uma hipótese composta, é possível derivar, de fato, um teste UMP, mas a maioria dos testes na prática, (por exemplo, \mathcal{H}_0 composta versus \mathcal{H}_A composta) não são, infelizmente, UMP.

Hipóteses Compostas: a função densidade de probabilidade (f.d.p.) do vetor das observáveis \mathbf{y} é $f_{\mathbf{y}}(\mathbf{y}|\mathbf{x})$ com $\mathbf{x} \in \Phi$, onde os parâmetros em \mathbf{x} podem pertencer à localização e forma da f.d.p. (por exemplo, média e variância). O espaço dos parâmetros Φ que contém todos os valores possíveis para \mathbf{x} é dividido em duas partes. Agora, as

hipóteses são dadas da seguinte forma:

$$\mathcal{H}_0 : \mathbf{x} \in \Phi_0; \text{ e } \mathcal{H}_A : \mathbf{x} \in \Phi \setminus \Phi_0 \quad (\text{B.2})$$

O conjunto $\Phi \setminus \Phi_0$ é um subconjunto de Φ e complementar a Φ_0 . Em outras palavras, $\Phi \setminus \Phi_0 = \{\mathbf{x} \in \Phi | \mathbf{x} \notin \Phi_0\}$. As hipóteses nulas e alternativas juntas cobrem todo o espaço dos parâmetros.

Teste da razão de verossimilhança generalizado (*Generalized likelihood ratio*): a f.d.p. do vetor das observáveis \mathbf{y} especificada por $f_{\mathbf{y}}(\mathbf{y}|\mathbf{x})$ é uma função de \mathbf{x} . Portanto, a especificação implica uma família inteira de f.d.p. Para o (dado) observado \mathbf{y} , considera-se $f_{\mathbf{y}}(\mathbf{y}|\mathbf{x})$ como uma função de \mathbf{x} . Isso é chamado de função de verossimilhança de \mathbf{y} . Quando \mathbf{x} varia, a função de verossimilhança produz as densidades de probabilidade de todas as f.d.p. possíveis para o vetor de amostra observado \mathbf{y} . Maximizamos a função de probabilidade por meio do método de estimação de verossimilhança (MLE). Isso vale para \mathbf{x} restrito ao conjunto Φ_0 sob a hipótese nula, bem como para o caso irrestrito $\mathbf{x} \in \Phi$. O teste da razão de verossimilhança generalizado (GLR) é definido como:

$$\text{rejeitar } \mathcal{H}_0 \text{ se } \frac{\max_{\mathbf{x} \in \Phi_0} f_{\mathbf{y}}(\mathbf{y}|\mathbf{x})}{\max_{\mathbf{x} \in \Phi} f_{\mathbf{y}}(\mathbf{y}|\mathbf{x})} < c, \quad \text{aceitar caso contrário} \quad (\text{B.3})$$

em que $c \in (0, 1)$.

O teste GLR em B produz uma decisão binária, ou seja, testa-se uma hipótese nula contra uma única hipótese alternativa. O numerador implica na maximização da probabilidade dentro do subconjunto $\Phi_0 \subset \Phi$ apresentado pela hipótese nula \mathcal{H}_0 . O denominador equivale a uma maximização sobre todo o conjunto do espaço dos parâmetros Φ . Quando o numerador é, de certa forma (especificado por c), menor que o denominador, a hipótese nula \mathcal{H}_0 provavelmente é falsa e, portanto, concluímos, com base na amostra, que existem evidências estatísticas suficientes para rejeitar a hipótese nula \mathcal{H}_0 .

O objetivo é tomar uma decisão correta para poder, eventualmente, calcular as estimativas para os parâmetros desconhecidos de interesse. Isso também nos fornece critérios como confiabilidade para controlar a qualidade de nossos estimadores finais.

Dois tipos de testes de hipóteses são geralmente empregados em geodésia: teste de observação (*observation test*) e o teste de significância de parâmetros (*parameters significance test*). Embora os testes sejam conduzidos em espaços diferentes, um no espaço das observações e outro no espaço dos parâmetros, ambos são equivalentes. Aqui, nós direcionamos os esforços para o teste no âmbito das observações.

Teste de Observação (*Observation testing*): aqui restringimos para o caso de observáveis com distribuição normal, embora algumas pesquisas já sinalizam a não normalidade dos dados em algumas aplicações geodésicas (LEHMANN, 2015; ROFATTO et al.,

2019). Nesta Tese, o modelo estocástico (modelo de incerteza) das observáveis não está sujeito a discussão ou decisão. Parte-se do pressuposto que o modelo estocástico é bem conhecido e definido. Neste caso, assumimos que o modelo de incerteza sob \mathcal{H}_0 seja dado pela distribuição normal multivariada, ou seja:

$$\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_e). \quad (\text{B.4})$$

em que $\mathbf{e} \in \mathbb{R}^{n \times 1}$ é o vetor de incertezas com suas componentes descritas na matriz covariância \mathbf{Q}_e .

A hipótese nula \mathcal{H}_0 , que também é chamada de hipótese de teste, corresponde a um modelo matemático supostamente válido que descreve os dados observados. Quando assumido como “verdadeiro”, métodos de estimação, normalmente na abordagem dos mínimos quadrados, são aplicados para estimar os parâmetros do modelo. Por outro lado, pode ser proposto um modelo alternativo, denotado por \mathcal{H}_A , quando houver dúvidas sobre o nível de confiabilidade do modelo em \mathcal{H}_0 . As duas hipóteses compostas sobre a esperança de \mathbf{y} são dadas como:

$$\mathcal{H}_0 : \mathbb{E}\{\mathbf{y}\} = \mathbf{A}\mathbf{x} + \mathbb{E}\{\mathbf{e}\} = \mathbf{A}\mathbf{x}; \quad \mathbb{D}\{\mathbf{y}\} = \mathbf{Q}_e \quad (\text{B.5})$$

versus

$$\mathcal{H}_A : \mathbb{E}\{\mathbf{y}\} = \mathbf{A}\mathbf{x} + \mathbf{C}\nabla + \mathbb{E}\{\mathbf{e}\} = \begin{pmatrix} \mathbf{A} & \mathbf{C} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \nabla \end{pmatrix} \quad (\text{B.6})$$

em que $\mathbb{E}\{\cdot\}$ é o operador de esperança matemática, $\mathbb{D}\{\cdot\}$ é o operador de dispersão, $\mathbf{y} \in \mathbb{R}^{n \times 1}$ é o vetor das observáveis, $\mathbf{A} \in \mathbb{R}^{n \times u}$ é a matriz *design* de posto completo u , $\mathbf{x} \in \mathbb{R}^{u \times 1}$ é o vetor dos parâmetros desconhecidos, $\mathbf{e} \in \mathbb{R}^{n \times 1}$ é o vetor das incertezas e $\mathbf{Q}_e \in \mathbb{R}^{n \times n}$ é matriz covariância (simétrica positiva-definida) das observáveis \mathbf{y} .

Na hipótese alternativa em (B.6), q parâmetros adicionais (desconhecidos) no vetor $\nabla \in \mathbb{R}^{q \times 1}$ são relacionados com a esperança de \mathbf{y} por meio da matriz $\mathbf{C} \in \mathbb{R}^{n \times q}$. A matriz \mathbf{C} é assumida ser de posto completo, ou característica completa, i.e. $\text{car}(\mathbf{C}) = q$. As colunas das matrizes \mathbf{A} e \mathbf{C} também são assumidas linearmente independentes, i.e. $\text{car}(\mathbf{A} \ \mathbf{C}) = u + q$. A matriz \mathbf{C} prescreve como os efeitos não modelados se traduzem nas observações individuais, ou seja, em todos os elementos do vetor \mathbf{y} .

O número de parâmetros que descreve a esperança das observáveis é estendido na hipótese alternativa \mathcal{H}_A . O objetivo é acomodar efeitos dos outliers no processo de medição. O teste da razão de verossimilhança generalizado (GLR) nos fornecerá uma decisão sobre se as variáveis explicativas adicionais em ∇ devem ou não ser levadas em consideração.

Estatística de teste: a f.d.p. sob \mathcal{H}_0 é $\mathcal{H}_0 : \mathbf{f}_y(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{A}\mathbf{x}, \mathbf{Q}_e)$, e sob $\mathcal{H}_A : \mathcal{N}(\mathbf{A}\mathbf{x} + \mathbf{C}\nabla, \mathbf{Q}_e)$. Neste caso, as duas partes do GLR são $\max_{\mathbf{x}} \mathcal{N}(\mathbf{A}\mathbf{x}, \mathbf{Q}_e)$ com $\mathbf{x} \in \mathbb{R}^{u \times 1}$, e $\max_{\mathbf{x}, \nabla} \mathcal{N}(\mathbf{A}\mathbf{x} + \mathbf{C}\nabla, \mathbf{Q}_e)$ com $\mathbf{x} \in \mathbb{R}^{u \times 1}$ e $\nabla \in \mathbb{R}^{q \times 1}$. Maximizando, essas

probabilidades levam aos estimadores de máxima verossimilhança (MLE); ou aqui (BLUE), pois é idêntico ao MLE para uma distribuição normal. Pode-se mostrar que a estatística de teste relacionada a esse problema é dada como:

$$\mathbf{T}_q = \hat{\mathbf{e}}_0^T \mathbf{Q}_e^{-1} \mathbf{C} (\mathbf{C}^T \mathbf{Q}_e^{-1} \mathbf{Q}_{\hat{\mathbf{e}}_0} \mathbf{Q}_e^{-1} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{Q}_e^{-1} \hat{\mathbf{e}}_0 \quad (\text{B.7})$$

em que $\hat{\mathbf{e}}_0$ é o vetor dos resíduos dos mínimos quadrados sob a hipótese nula \mathcal{H}_0 .

A estatística de teste \mathbf{T}_q é distribuída da seguinte maneira:

$$\mathcal{H}_0 : \mathbf{T}_q \sim \chi^2(q, 0), \quad \mathcal{H}_A : \mathbf{T}_q \sim \chi^2(q, \lambda) \quad (\text{B.8})$$

em que o parâmetro de não-centralidade λ é dado como:

$$\lambda = \nabla^T \mathbf{C}^T P_A^{\perp T} \mathbf{Q}_e^{-1} P_A^{\perp T} \mathbf{C} \nabla = \|\mathbf{P}_A^{\perp} \mathbf{C} \nabla\|_{\mathbf{Q}_e^{-1}}^2 \quad (\text{B.9})$$

cujo resultado é igual a norma quadrática do vetor $\mathbf{P}_A^{\perp} \mathbf{C} \nabla$. O teste GLR é, portanto, usado para decidir entre o modelo estendido sob a hipótese alternativa \mathcal{H}_A e o modelo padrão sob a hipótese nula \mathcal{H}_0 .

Casos Especiais: o número de parâmetros no vetor ∇ da hipótese alternativa em (B.6) pode variar de 1 a $n - u$, i.e. $1 \leq q \leq n - u$. Quando o número de parâmetros adicionais é igual à redundância do sistema, i.e. $q = n - u$, um teste global do model (*overall model test*) é realizado. Neste caso extremo, a redundância da hipótese alternativa é $r = n - (u + q) = n - (u + n - u) = 0$. Logo, não há necessidade de especificar qualquer hipótese alternativa específica. Outro caso especial é quando tomamos $q = 1$, ou seja, estendemos o modelo da hipótese nula com apenas um parâmetro adicional. Neste caso, tem-se um teste com natureza local (*local model test*), pois cada observação individual pode ser testada.

Teste Global do modelo: no caso limitante $q = n - u$, não há redundância na hipótese alternativa em (2.24). Isso implica imediatamente que $\hat{\mathbf{e}}_A = 0$, onde $\hat{\mathbf{e}}_A$ é vetor dos resíduos dos mínimos quadrados sob a hipótese alternativa. Neste caso, pode-se mostrar que a estatística de teste (B.7) produz:

$$\mathbf{T}_{q=n-u} = \|\hat{\mathbf{e}}\|_{\mathbf{Q}_e^{-1}}^2 = \hat{\mathbf{e}}_0^T \mathbf{Q}_e^{-1} \hat{\mathbf{e}}_0, \quad (\text{B.10})$$

que é expressa como a forma quadrática (norma ao quadrado) dos resíduos. Esse teste é conhecido na literatura nacional como Teste Global do Ajustamento. Observe que tanto \mathbf{A} quanto \mathbf{C} devem ser de característica, ou posto, completo e, portanto, a matriz $(\mathbf{A} \ \mathbf{C})$ deve ser quadrada e invertível. De acordo com a hipótese alternativa \mathcal{H}_A , é permitido que o vetor de observações \mathbf{y} esteja em qualquer lugar no espaço das observações \mathbb{R}^n , isto é, $\mathbf{y} \in \mathbb{R}^n$. Portanto, não é possível ainda ter em mente uma *assinatura* de erro específica.

Distribuição: o valor da estatística de teste em (B.10) é igual à norma do vetor $\hat{\mathbf{e}}_0$ e fornece uma medida geral do grau de inconsistência do modelo. A estatística de teste $\mathbf{T}_{q=n-u}$ é distribuída da seguinte forma:

$$\mathcal{H}_0 : \mathbf{T}_{q=n-u} \sim \chi^2(n-u, 0); \quad \mathcal{H}_A : \mathbf{T}_{q=n-u} \sim \chi^2(n-u, \lambda) \quad (\text{B.11})$$

em que o parâmetro de não-centralidade λ na hipótese alternativa é mencionada em (B.9).

Variância do peso unitário: quando a matriz covariância das observáveis é decomposta como $\mathbf{Q}_e = \sigma_0^2 \mathbf{Q}$, com $\mathbf{Q} \in \mathbb{R}^{n \times n}$ a matriz cofatora, então

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{e}}_0^T \mathbf{Q}^{-1} \hat{\mathbf{e}}_0}{n-u} \quad (\text{B.12})$$

é um estimador não-tendencioso (*unbiased*) para a variância do peso unitário σ_0^2 , pois $\mathbb{E}\{\hat{\mathbf{e}}_0^T \mathbf{Q}^{-1} \hat{\mathbf{e}}_0\} = \sigma_0^2(n-u)$. A relação de (B.12) com a estatística do teste global do modelo (Teste Global do Ajustamento) é dada por:

$$\frac{\hat{\sigma}^2}{\sigma_0^2} = \frac{\hat{\mathbf{e}}_0^T \mathbf{Q}^{-1} \hat{\mathbf{e}}_0}{\sigma_0^2(n-u)} = \frac{\mathbf{T}_{q=n-u}}{n-u} \quad (\text{B.13})$$

Observe que $\frac{\hat{\sigma}^2}{\sigma_0^2}$ coincide com $\frac{\mathbf{T}_{q=n-u}}{n-u}$, sendo que, este último, também pode ser tomado como uma estimador para o fator de variância do peso unitário.

A relação em (B.13) indica como aumentar ou diminuir o fator de variância a priori (σ_0^2) para atender a redundância do sistema em jogo. Um modelo de incerteza incorreto, descrito pela matriz de covariância das observáveis \mathbf{Q}_e , também pode causar a rejeição da hipótese nula \mathcal{H}_0 no teste global do modelo. Quando, por exemplo, os elementos da matriz \mathbf{Q}_e são considerados muito pequenos, o valor para o teste geral do modelo se torna muito grande. Portanto, é (mais) provável que a hipótese nula \mathcal{H}_0 seja rejeitada. Nesse caso, a estimativa obtida em (B.12) mede como escalar a variância a priori σ_0^2 .

Com apenas o teste global do modelo por meio da estatística de teste (B.10), não podemos descrever as razões que levaram à uma possível rejeição da hipótese nula. Em outras palavras, somos apenas capazes de concluir que o modelo dado na hipótese nula \mathcal{H}_0 não é válido. Efeitos adversos na coleta de dados, um modelo funcional (padrão) inadequado para os dados observados e/ou uma especificação inadequada das características do ruído das observáveis no modelo estocástico são exemplos de acontecimentos que podem levar à rejeição da hipótese nula. O teste global do modelo por si só não pode fornecer uma resposta.

Importante ressaltar que os estimadores (BLUE) não são afetados nos casos em que o fator de variância a priori σ_0^2 é desconhecido. Aqui, consideramos que o modelo de

incertezas é completamente conhecido, ou seja, todas as componentes da matriz covariância das observáveis (\mathbf{Q}_e) são bem definidas e especificadas.

Teste Local do modelo: no procedimento *Data-Snooping* proposto originalmente por Baarda (1968), o teste local do modelo (*local model test*) é aplicado com o objetivo de localizar um possível outlier nos dados observados. O teste é unidimensional, pois considera-se o limite inferior de $q = 1$ e, portanto, o número de parâmetros adicionais no vetor ∇ da hipótese alternativa é $q = 1$. Neste caso especial, cada observação é testada individualmente quanto ao outlier, o que implica que executaremos um número n de testes. Portanto, quando $q = 1$, a matriz \mathbf{C} reduz a um vetor \mathbf{c}_i e ∇ em um escalar ∇_i . Introduzindo esse vetor \mathbf{c}_i em (B.7), tem-se:

$$\mathbf{T}_{q=1} = \frac{(\mathbf{c}_i^T \mathbf{Q}_e^{-1} \hat{\mathbf{e}}_0)^2}{\mathbf{c}_i^T \mathbf{Q}_e^{-1} \mathbf{Q}_{\hat{\mathbf{e}}_0} \mathbf{Q}_e^{-1} \mathbf{c}_i} \quad (\text{B.14})$$

em que o numerador e denominador são ambas quantidades escalares. A estatística de teste $\mathbf{T}_{q=1}$ é distribuída como:

$$\mathcal{H}_0 : \mathbf{T}_{q=1} \sim \chi^2(1, 0); \quad \mathcal{H}_A : \mathbf{T}_{q=1} \sim \chi^2(1, \lambda) \quad (\text{B.15})$$

onde o parâmetro de não-centralidade λ pode ser obtido de acordo com a Equação B.9 para o caso $q = 1$.

Na pratica, é mais conveniente usar a raiz quadrada da estatística de teste em (B.14), conhecida como *w-test*, i.e.:

$$\mathbf{w}_i = \frac{\mathbf{c}_i^T \mathbf{Q}_e^{-1} \hat{\mathbf{e}}}{\sqrt{\mathbf{c}_i^T \mathbf{Q}_e^{-1} \mathbf{Q}_{\hat{\mathbf{e}}} \mathbf{Q}_e^{-1} \mathbf{c}_i}}, \forall i = 1, \dots, n \quad (\text{B.16})$$

A estatística de teste em (B.16) segue distribuição normal. Neste caso, tem-se:

$$\mathcal{H}_0 : \mathbf{w}_i \sim \mathcal{N}(0, 1); \quad \mathcal{H}_A : \mathbf{w}_i \sim \mathcal{N}(\nabla w_i, 1) \quad (\text{B.17})$$

com $\nabla w_i = (\mathbf{c}_i^T \mathbf{Q}_e^{-1} \mathbf{Q}_{\hat{\mathbf{e}}} \mathbf{Q}_e^{-1} \mathbf{c}_i)^{1/2} \nabla_i$. O parâmetro de não-centralidade está relacionado com ∇w_i da seguinte forma $\lambda = \nabla w_i^2$.

Data-Snooping: Uma aplicação importante do teste *w-test* é a *Detecção de Outliers*. Considera-se que um outlier afeta somente uma única observação. Para rastrear as observações, a fim de identificar o outlier, são formuladas n hipóteses alternativas. Todas elas são testadas individualmente contra o modelo padrão em \mathcal{H}_0 . Neste caso, o vetor \mathbf{c}_i consiste de elementos de 0 e 1, em que 1 significa a posição da observação a ser testada quanto a um possível outlier, enquanto 0 é atribuído para as demais observações. Por exemplo, se deseja testar uma observação na i -ésima posição, tem-se $\mathbf{c}_i = (0 \ 0 \ 0 \ \dots \ 1^{i\text{th}} \ 0 \ \dots \ 0)^T$. Como se tem n hipóteses alternativas, testa-se cada uma individualmente por meio da

estatística de teste na Equação (B.17), variando a posição i do elemento “1” de 1 a n no vetor \mathbf{c}_i . Esse procedimento é chamado de *Data-Snooping*.

O *Data-Snooping* é aplicado quando se tem conhecimento do fator de variância a priori σ^2 . Quando ele não é conhecido, é possível estimá-lo por meio da Equação (B.12). Neste caso, a distribuição da estatística de teste em (B.16) segue uma distribuição t-Student. Entretanto, neste caso, o teste global do ajustamento não é realizado e o procedimento *Data-Snooping* deve ser modificado utilizando uma distribuição pouco conhecida, denominada de distribuição Tau. Como o fator de variância a posteriori, estimado em função dos resíduos, pode ser afetado pela presença de outliers, o teste baseado na distribuição Tau é menos sensível à identificação de outliers que o *Data-Snooping* original.

Destacamos, que embora sejam formuladas n hipóteses alternativas para o procedimento *Data-Snooping* original, a regra de decisão é binária, ou seja, testa-se uma hipótese nula contra uma única hipótese alternativa. Na prática, como demonstrado, o *Data-Snooping* envolve múltiplas hipóteses alternativas, com múltiplas adaptações.

Resíduo Normalizado: Se a matriz covariância \mathbf{Q}_e das observáveis for diagonal, a expressão para a estatística w -test reduz para uma forma muito simples, ou seja:

$$w_i = \frac{\hat{\epsilon}_i}{\sigma_{\hat{\epsilon}_i}} \quad (\text{B.18})$$

com $\sigma_{\hat{\epsilon}_i} = (\mathbf{Q}_{\hat{\epsilon}_0})_{ii}^{1/2}$ o desvio-padrão do resíduo dos mínimos quadrados para i, \dots, n .