

**MINISTÉRIO DA DEFESA
EXÉRCITO BRASILEIRO
DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA
INSTITUTO MILITAR DE ENGENHARIA
CURSO DE MESTRADO EM ENGENHARIA CARTOGRÁFICA**

STEFANO SAMPAIO SURACI

**APLICAÇÃO DAS NORMAS L_1 E L_∞ EM REDES ALTIMÉTRICAS:
IDENTIFICAÇÃO DE *OUTLIERS* E CONSTRUÇÃO DO MODELO
ESTOCÁSTICO**

Rio de Janeiro

2018

INSTITUTO MILITAR DE ENGENHARIA

STEFANO SAMPAIO SURACI

**APLICAÇÃO DAS NORMAS L_1 E L_∞ EM REDES ALTIMÉTRICAS:
IDENTIFICAÇÃO DE *OUTLIERS* E CONSTRUÇÃO DO MODELO
ESTOCÁSTICO**

Dissertação de Mestrado apresentada ao Curso de Mestrado em Engenharia Cartográfica do Instituto Militar de Engenharia como requisito parcial para obtenção do título de Mestre em Ciências em Engenharia Cartográfica.

Orientador: Leonardo Castro de Oliveira, DE.

Rio de Janeiro

2018

c2018

INSTITUTO MILITAR DE ENGENHARIA

Praça General Tibúrcio, 80 – Praia Vermelha

Rio de Janeiro – RJ CEP: 22290-270

Este exemplar é de propriedade do Instituto Militar de Engenharia, que poderá incluí-lo em base de dados, armazenar em computador, microfilmар ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es) e do(s) orientador(es).

526 Suraci, Stefano Sampaio

S961a Aplicação das normas L1 e Loo em redes altimétricas: identificação de outliers e construção do modelo estocástico / Stefano Sampaio Suraci; orientado por Leonardo Castro de Oliveira – Rio de Janeiro: Instituto Militar de Engenharia, 2018.

90p.: il.

Dissertação (Mestrado) – Instituto Militar de Engenharia, Rio de Janeiro, 2018.

1. Curso de Engenharia Cartográfica – teses e dissertações. 2. Redes altimétricas. 3. Modelo estocástico. I. Oliveira, Leonardo Castro de. II. Instituto Militar de Engenharia. III. Título.

INSTITUTO MILITAR DE ENGENHARIA

STEFANO SAMPAIO SURACI

**APLICAÇÃO DAS NORMAS L_1 E L_∞ EM REDES ALTIMÉTRICAS:
IDENTIFICAÇÃO DE *OUTLIERS* E CONSTRUÇÃO DO MODELO ESTOCÁSTICO**

Dissertação de Mestrado apresentada ao Curso de Mestrado em Engenharia Cartográfica do Instituto Militar de Engenharia, como requisito parcial para obtenção do título de Mestre em Ciências em Engenharia Cartográfica.

Orientador: Leonardo Castro de Oliveira, DE.

Aprovada em 19 de dezembro de 2018 pela seguinte banca examinadora:

Prof. Leonardo Castro de Oliveira, D.E. do IME – Presidente

Prof. Ivandro Klein, D.E. do IFSC

Eng. Roberto Teixeira Luz, D.C. do IBGE

Profa. Heloísa Alves Silva Marques, D.C. do IME

Rio de Janeiro

2018

À minha bisavó Aida,
Às minhas avós Cremilda e Adelina,
Aos meus pais Sonia e Domenico,
Ao meu padrasto Mauri,
À minha esposa Patricia Loh,
À minha irmã Patricia Suraci,
Ao meu amigo Gil,
À minha madrinha Isabel,
Em especial, à nova e à novíssima geração:
Meu irmão Luiz Claudio, e
Minha filha Sabrina, respectivamente.

AGRADECIMENTOS

Ao meu orientador Prof. Dr. Leonardo Castro de Oliveira, idealizador dessa pesquisa. Pela confiança depositada, pela orientação, lealdade e amizade.

À Profa. Dra. Heloísa e ao Prof. Dr. Haroldo, pelas valiosas sugestões apresentadas na minha Defesa da Proposta de Pesquisa.

Ao Prof. Dr. Ivandro Klein, ao Prof. Dr. Roberto Luz e à Prof. Dra. Heloísa, integrantes da banca examinadora, pelas relevantes contribuições para o trabalho.

Aos demais professores da Seção de Engenharia Cartográfica, pelas sugestões apresentadas para a pesquisa nos Seminários por mim apresentados.

À Seção de Ensino de Engenharia de Computação, pelo atendimento às demandas de utilização das máquinas virtuais de alto desempenho do IME para condução de experimentos. Em especial, ao Maj Vasconcelos, por viabilizar esse suporte fundamental para a pesquisa.

À Profa. Dra. Raquel, ao Maj Roberto Gomes e ao S Ten Aguiar, pelo apoio com questões administrativas durante o curso.

Aos companheiros de pós-graduação, da própria Engenharia Cartográfica ou de outros cursos do IME, pela camaradagem e momentos de descontração: Maj Gustavo, Maj Heber, Maj Menezes, Cap Kitagawa, 1º Ten Raphael, 1º Ten Dave, Leonardo, Verônica e Henrique.

À minha esposa Patricia, pelo companheirismo e estímulo na busca do que me faz feliz.

À minha mãe Sonia, pelo esforço para que eu tivesse uma educação básica de qualidade.

Ao meu padrasto Mauri, à minha mãe Sonia e à minha esposa Patricia, pela ajuda em momentos difíceis de minha carreira militar.

Ao Exército Brasileiro, por mais essa oportunidade de formação, aprendizado e superação.

Ao Instituto Brasileiro de Geografia e Estatística, por dispor parte dos dados utilizados nos experimentos dessa dissertação livremente em seu sítio na *internet*.

SUMÁRIO

LISTA DE ILUSTRAÇÕES	8
LISTA DE TABELAS	9
LISTA DE ABREVIATURAS E SÍMBOLOS.....	10
LISTA DE SIGLAS	11
1 INTRODUÇÃO	14
1.1 Problemas	14
1.2 Hipóteses e Objetivo	16
1.3 Justificativa.....	17
1.4 Organização da Dissertação	17
2 NORMAS L_1 E L_∞ NO AJUSTAMENTO DAS OBSERVAÇÕES	19
2.1 Características das aplicação das normas L_1 e L_∞	19
2.2 Formulação em redes geodésicas	22
2.2.1 Formulação da ML_1	23
2.2.2 Formulação da ML_∞	24
3 CONTEXTO DA APLICAÇÃO DA NORMA L_1	26
3.1 <i>Outliers</i> x erros grosseiros.....	26
3.1.1 Considerações sobre <i>outliers</i> de forma objetiva	29
3.2 Tratamento de <i>outliers</i> em redes geodésicas	32
3.2.1 O <i>data snooping</i>	35
3.2.2 Identificação de <i>outliers</i> pela ML_1 e seu estado da arte	36
3.3 Influência da geometria da rede	38
3.4 Avaliação dos classificadores de <i>outliers</i>	41
3.4.1 Simulação de Monte Carlo aplicada.....	42
4 EXTENSÃO DO CONCEITO DE ESTIMAÇÃO ROBUSTA E CONTEXTO DA APLICAÇÃO DA NORMA L_∞ EM REDES ALTIMÉTRICAS.....	44
4.1 Conceitos relacionados ao levantamento altimétrico	45
4.2 Aspectos relevantes para a estimação robusta.....	47
4.3 Modelo estocástico usual em redes altimétricas	48
4.4 Uma abordagem alternativa para o modelo estocástico	51

5	PROCEDIMENTOS PROPOSTOS E MÉTODOS DE AVALIAÇÃO.....	54
5.1	Visão geral.....	54
5.2	Alternativa de Pesos para a ML1 e de Parâmetro de Comparação	56
5.3	O Classificador VL1 proposto.....	57
5.3.1	Obtenção de um VC válido para qualquer rede	59
5.4	Alternativa de Pesos para o MMQ	60
6	EXPERIMENTOS E ANÁLISE DE RESULTADOS	62
6.1	Redes	62
6.1.1	Rede Matemática – base para SMC	62
6.1.2	Rede Real – trecho da RAAP	63
6.2	Simulações a partir da Rede Matemática	64
6.2.1	Simulação 1 – cenários sem <i>outliers</i>	64
6.2.2	Simulação 2 – cenários com <i>outliers</i>	65
6.3	Experimento 1 - Alternativa de Pesos para a ML1 e de Parâmetro de Comparação	67
6.4	Experimentos referentes ao Classificador VL1	68
6.4.1	Experimento 2	68
6.4.2	Experimento 3	69
6.4.3	Experimento 4	72
6.4.4	Experimento 5	74
6.5	Experimento 6 - Alternativa de Pesos para o MMQ	75
7	CONSIDERAÇÕES FINAIS	78
7.1	Conclusões	78
7.2	Sugestões de trabalhos futuros	80
8	REFERÊNCIAS	82
9	APÊNDICES.....	86
9.1	Apêndice 1: Observações da Rede Real.....	86
9.2	Apêndice 2: Memorial dos principais VCs testados – Experimento 3.....	88
9.3	Apêndice 3: Memorial dos principais VCs testados – Experimento 5.....	90

LISTA DE ILUSTRAÇÕES

FIG. 1.1	Exemplo da deficiência de robustez do MMQ e da ML^∞	15
FIG. 3.1	Posicionamento de <i>outliers</i> dentre as observações	28
FIG. 3.2	Erro aleatório discrepante	31
FIG. 3.3	Erro total não discrepante, apesar de erro grosseiro acima do limite	31
FIG. 3.4	Fases do tratamento de <i>outliers</i>	32
FIG. 3.5	Redes de nivelamento com e sem geometria confiável contra 1 <i>outlier</i>	39
FIG. 3.6	Rede analisada contra 2 <i>outliers</i>	41
FIG. 4.1	Interpretações para a estimação robusta de redes	44
FIG. 4.2	Lance de nivelamento geométrico	46
FIG. 4.3	Exemplos de componentes de uma rede altimétrica	46
FIG. 5.1	Contribuições dessa pesquisa em um possível processo de estimação robusta	54
FIG. 5.2	Aplicação do Classificador VL1	58
FIG. 5.3	Diagrama de atividades para construção do VCL1	58
FIG. 5.4	Construção do modelo estocástico alternativo proposto	60
FIG. 6.1	Geometria da Rede Matemática	63
FIG. 6.2	Trecho da RAAP utilizado nos experimentos	64
FIG. 6.3	Sequência de obtenção dos cenários de rede com 1 <i>outlier</i> – Simulação 2	66
FIG. 6.4	VCs (mm) x quantidade de acertos – escolha do $VCL1_{ V }$	70
FIG. 6.5	Tempo médio de execução (em segundos) – Experimento 4	73

LISTA DE TABELAS

TAB. 2.1	Outras características da minimização das normas vetoriais de interesse	21
TAB. 3.1	Equações medianas das observações da rede da FIG 3.5a.....	39
TAB. 3.2	Equações medianas das observações da rede da FIG 3.5b.....	39
TAB. 3.3	Equações medianas estendidas	41
TAB. 4.1	Conjecturas do modelo estocástico usual.....	50
TAB. 4.2	Máximo resíduo absoluto no ajustamento da FIG 1.1	52
TAB. 4.3	Modelo estocástico da ML_{∞} x minimização de MRAP e MRA.....	53
TAB. 6.1	Desníveis exatos – Rede Matemática.....	63
TAB. 6.2	Valores de LimI e LimF para cada tipo de <i>outlier</i>	67
TAB. 6.3	PPCOMR – cenários com 1 <i>outlier</i> – Experimento 1	67
TAB. 6.4	PPCOMR – cenários com 2 <i>outliers</i> – Experimento 1.....	68
TAB. 6.5	Observações da rede – Experimento 2	69
TAB. 6.6	Resíduos absolutos – Experimento 2	69
TAB. 6.7	TS – 1 <i>outlier</i> – Experimento 3.....	70
TAB. 6.8	TS – 2 <i>outliers</i> – Experimento 3	71
TAB. 6.9	TS – 2 <i>outliers</i> controláveis – Experimento 3.....	71
TAB. 6.10	TS – 2 <i>outliers</i> não controláveis – Experimento 3.....	71
TAB. 6.11	Sucesso na identificação de <i>outliers</i> – Experimento 4.....	73
TAB. 6.12	TS – 1 <i>outlier</i> – Experimento 5.....	74
TAB. 6.13	Quantidade de resultados mais desejáveis – Experimento 6 – Simulação 1	75
TAB. 6.14	Resultados do ajustamento pelo MMQ – Experimento 6 – Rede Real.....	76

LISTA DE ABREVIATURAS E SÍMBOLOS

ABREVIATURAS

ME usual	-	Modelo estocástico usual
ME proposto	-	Modelo estocástico proposto

SÍMBOLOS

σ	-	desvio padrão
μ	-	média
α	-	significância
u.m.	-	unidade de medida
VCL1	-	valor de corte do Classificador VL1
DP _x	-	Vetor das raízes quadradas dos valores absolutos dos elementos da diagonal principal da MVC dos parâmetros estimados
DP _v	-	Vetor das raízes quadradas dos valores absolutos dos elementos da diagonal principal da MVC dos resíduos

LISTA DE SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
DP	Desvio padrão
DS	<i>Data Snooping</i>
EM	Equações medianas
EPI	Estratégia de Pesos Iterativos
IAG	<i>International Association of Geodesy</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IHRS	<i>International Height Reference System</i>
MEM	Método das Equações Medianas
ML1	Minimização da norma L1
ML ∞	Minimização da norma L ∞
MMQ	Método dos Mínimos Quadrados
MRA	Máximo Resíduo Absoluto
MRAP	Máximo Resíduo Absoluto Ponderado
MVC	Matriz Variância Covariância
MVI	Método das Visadas Iguais
PPCOMR	Percentual de ocorrências do parâmetro de comparação do <i>outlier</i> como sendo o maior dentre as observações da rede
RAAP	Rede Altimétrica de Alta Precisão (do Sistema Geodésico Brasileiro)
RRNN	Referências de Nível
SMC	Simulação de Monte Carlo
TS	Taxa de Sucesso
VC	Valor de Corte

RESUMO

Essa dissertação investigou a aplicação da minimização da norma L_1 (ML_1) e da norma L_∞ (ML_∞) no contexto da estimação robusta de redes altimétricas, mais especificamente nas etapas de identificação de *outliers* e na definição do modelo estocástico para ajustamento final da rede pelo MMQ, respectivamente. Discussões teóricas sobre os conceitos de *outlier* e estimação robusta também foram apresentadas.

Nos experimentos, redes de nivelamento simuladas pela Técnica de Monte Carlo e dados reais de nivelamento geométrico foram empregados. Para testes relativos à identificação de *outliers* pela ML_1 , *outliers* foram propositalmente inseridos nas redes analisadas. Nessas, foi verificado que o ajustamento pela ML_1 com pesos unitários apresentou condições vantajosas para a identificação de *outliers* em relação ao modelo estocástico usual de pesos proporcionais ao inverso do comprimento das linhas de nivelamento, seja esse último considerando resíduo absoluto ou resíduo absoluto ponderado como parâmetro para identificação de *outliers*.

Critérios objetivos para identificação de *outliers* caracterizados por Valores de Corte para os resultados do ajustamento pela ML_1 a partir dos quais a respectiva observação passa a ser classificada como *outlier* foram propostos. A taxa de sucesso na identificação de *outliers* desses critérios objetivos foi superior à do procedimento *data snooping* em alguns cenários com *outliers*, especialmente para os casos de geometria da rede deficiente. O tempo de execução computacional da identificação de *outliers* pela ML_1 também foi menor que o do *data snooping*.

A aplicação da ML_∞ proposta é posterior ao tratamento de *outliers*. Assim, os testes foram realizados em redes simuladas e dados reais sem *outliers*. Um modelo estocástico alternativo para ajustamento final da rede pelo MMQ que aproveitou a característica de minimização do máximo resíduo absoluto da rede no ajustamento pela ML_∞ com pesos iguais foi proposto. Além dessa minimização, o ajustamento da rede pelo MMQ com o modelo proposto gerou, na maioria dos casos, resíduos e precisão desses e dos parâmetros estimados mais homogêneos, com menor desvio padrão, que aqueles com a abordagem usual para o modelo estocástico.

ABSTRACT

This dissertation investigated the application of L1 norm minimization (ML1) and L^∞ norm minimization (ML^∞) in the context of the robust estimation of leveling networks, more specifically in the stages of identification of outliers and defining the stochastic model for final network adjustment by the Least Squares method, respectively. Theoretical discussions on outlier concepts and robust estimation were also presented.

In the experiments, leveling networks simulated by the Monte Carlo technique and real leveling data were used. For tests related to the identification of outliers by ML1, outliers were purposely inserted in the analyzed networks. It was verified that the adjustment by the ML1 with unitary weights presented advantageous conditions to the identification of outliers in relation to the usual stochastic model of weights proportional to the inverse of the length of respective leveling lines, be it the latter considering absolute residual or weighted absolute residue as parameter for identification of outliers.

Objective criteria for identifying outliers characterized by cut-off values for the results of the ML1 adjustment from which the respective observation becomes classified as outlier were proposed. The success rate in identifying outliers of these objective criteria was higher than that of the *data snooping* procedure in some scenarios with outliers, especially for cases of poor network geometry. The computational execution time of the outliers identification by the ML1 was shorter in relation to the *data snooping*.

The application proposed for ML^∞ is after the treatment of outliers. Thus, the tests were performed on simulated networks and real data without outliers. An alternative stochastic model for final network adjustment by Least Squares that took advantage of the characteristic of minimization of the maximum absolute residue of the network in the adjustment by the ML^∞ with equal weights was proposed. Besides this minimization, the adjustment of the network by Least Squares with the proposed model generated, in the majority of cases, residuals and precision of these and of the estimated parameters more homogeneous, with lower standard deviation, than those with the usual approach for the stochastic model.

1 INTRODUÇÃO

1.1 PROBLEMAS

Na estimação de redes geodésicas, o número de observações é superior ao de incógnitas. Esta é a base para indicação da acurácia e da confiabilidade de redes geodésicas por meio de redundância. Os inevitáveis erros de medida tornam o sistema inconsistente (GEMAEL *et al*, 2015). O ajustamento da rede é usualmente realizado pelo Método dos Mínimos Quadrados (MMQ).

O MMQ minimiza a soma dos quadrados ponderados dos resíduos (diferença entre valores ajustados e observados) e seus resultados são os de mínima variância para os parâmetros estimados e de máxima verossimilhança, essa sob a premissa de ocorrência de somente erros aleatórios normalmente distribuídos nas observações (GHILANI, 2010). Porém, trata-se de um método pouco robusto, ou seja, sensível a pequenos desvios, como uma pequena quantidade de *outliers* nas observações, em relação às suas premissas (BUSTOS, 1981).

Uma possível definição para *outlier* é a de uma observação com valor atípico em relação às demais da amostra. É comum que geodestas erroneamente entendam que *outlier* é sinônimo de erro grosseiro. Uma análise mais detalhada para o conceito de *outlier* será apresentada na seção 3.1. A FIG 1.1 ilustra o ajustamento a uma reta horizontal dos pontos (0,0), (1,0), (2,5), (3,0) e (4,0), todos com pesos unitários e com as coordenadas em determinada unidade de medida (u.m.). Nesse contexto, a terceira observação (ponto) pode ser considerada um *outlier*.

No caso dos ajustes pelo MMQ e pela Minimização da Norma de L^∞ (ML^∞), ou Minimização da Norma de *Chebyshev*, o *outlier* distribuiu resíduo por todas as demais “boas” observações, sendo que a ML^∞ distribuiu ainda mais que o MMQ. Isso não aconteceu no ajustamento pela Minimização da Norma L_1 (ML_1), ou Método dos Valores Absolutos, que é mais robusto e, assim, mais resistente a *outliers*.

Devido ao efeito indesejável dos *outliers*, procedimentos e métodos que dão o devido tratamento a eles, diminuindo ou retirando sua influência nos parâmetros estimados, são empregados na estimação de redes geodésicas. Para a identificação de *outliers* da rede, o procedimento *data-snooping* (DS), proposto inicialmente por BAARDA (1968), é o mais usual. Porém, ele é desenvolvido a partir do ajustamento pelo MMQ, cuja distribuição de

resíduos no ajustamento pode, entre outros problemas, ocultar os próprios *outliers* existentes, efeito chamado de mascaramento.

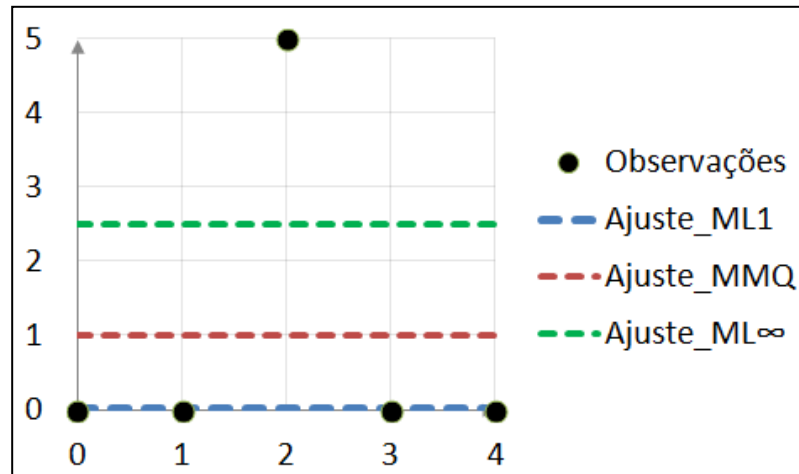


FIG. 1.1: Exemplo da deficiência de robustez do MMQ e da ML^∞ .

De fato, nenhum dos procedimentos conhecidos para identificação de *outliers* em redes geodésicas garante 100% de sucesso nessa tarefa. Um dos demais que aparece na literatura corresponde à realização de um “pré-ajuste” da rede pela ML1, visando à identificação de *outliers* a partir dos resíduos absolutos desse “pré-ajuste” (o ajustamento propriamente dito da rede para estimação de seus parâmetros é feito via MMQ, após ação corretiva sobre os *outliers* previamente identificados pela ML1). Nessa pesquisa, algumas questões inéditas relativas ao ajustamento pela ML1 na identificação de *outliers* em redes altimétricas foram investigadas. O estado da arte da ML1 e a descrição dessas questões foram apresentados na seção 3.2.2.

Por outro lado, o modelo estocástico usual em redes altimétricas adota pesos das observações como inversamente proporcionais ao comprimento das respectivas linhas de nivelamento. É baseado em conjecturas cuja garantia de ocorrência na prática dos levantamentos é pouco viável, como a utilização de equipamentos com a mesma precisão e adoção da mesma distância em todas as visadas da rede, o que justifica a busca por abordagens alternativas para o modelo estocástico de redes altimétricas. Uma análise das conjecturas do modelo estocástico usual foi feita na seção 4.3.

Ainda sobre o exemplo da FIG 1.1, ressalta-se que o máximo resíduo absoluto (MRA) do ajustamento pela ML^∞ foi o menor dentre os três métodos. Ademais, no contexto da estimação de uma rede, após o tratamento de *outliers*, assim como no caso do MMQ, a deficiência de resistência a *outliers* da ML^∞ deixa de ser um problema. Assim, em cenários

nos quais todas as observações são confiáveis, nesse contexto, entendido como isentas de *outliers*, essa minimização do MRA no ajustamento, que significa uma mínima distorção máxima entre as observações coletadas e ajustadas, pode ser aproveitada.

Nessa pesquisa, essa propriedade do resultado do ajustamento pela ML^∞ foi aproveitada para definição de um modelo estocástico alternativo para o ajustamento final da rede (livre de *outliers*) pelo MMQ. Ou seja, trata-se de mais um “pré-ajuste” da rede (similar ao que já é feito com o “pré-ajuste” pela ML_1 para identificação de *outliers*), mas no caso da ML^∞ com o objetivo de construir um modelo estocástico alternativo para ela.

Apesar da presente pesquisa tratar de redes altimétricas, as investigações realizadas podem ser adaptadas para outros tipos de redes geodésicas. Entretanto, com este recorte realizado, a fundamentação teórica, os experimentos e os resultados da pesquisa foram especificamente dedicados à altimetria, que tem recebido especial atenção atualmente no contexto dos sistemas de referência.

Prova disto é a Resolução 1/2015 (IAG, 2015) da *International Association of Geodesy* (IAG), que definiu o *International Height Reference System* (IHRs), com geoide único e adoção de números geopotenciais, reconhecendo a necessidade de um sistema global de alta acurácia para investigação de mudanças na Terra, como a elevação dos níveis dos oceanos.

1.2 HIPÓTESES E OBJETIVO

A pesquisa está associada a três hipóteses inéditas, a saber:

I. Em redes altimétricas, a adoção de pesos unitários para todas as observações no ajustamento pela ML_1 apresenta condições vantajosas para a identificação de *outliers* em relação ao modelo estocástico usual de pesos proporcionais ao inverso do comprimento das respectivas linhas de nivelamento.

II. Há cenários de redes altimétricas em que a ML_1 , associada a um critério objetivo adequado, é mais eficaz que o DS na identificação de *outliers*.

III. O ajustamento pelo MMQ de redes altimétricas livres de *outliers* com modelagem estocástica advinda do MRA gera resultados mais homogêneos que com o modelo usual.

Assim sendo, a pesquisa teve como objetivo investigar aplicações da minimização das normas L_1 e L^∞ no contexto da estimação robusta de redes altimétricas, mais especificamente nas etapas de identificação de *outliers* e na definição do modelo estocástico para ajustamento final da rede pelo MMQ, respectivamente.

1.3 JUSTIFICATIVA

A demanda por coordenadas cada vez mais acuradas e confiáveis é crescente na atualidade, para os diversos fins. No caso das coordenadas altimétricas, suas aplicações são inúmeras: obras desenvolvimentistas de construção de barragens e dutos; monitoramento da Terra, que culminou com a definição do IHRS pela IAG; prevenção de inundações e deslizamento de encostas habitadas, etc. O conhecimento do relevo é fundamental também para o planejamento de operações militares. As variações altimétricas, por exemplo, em conjunto com outros fatores, determinam a viabilidade e a velocidade do deslocamento das tropas e equipamentos.

Entretanto, a qualidade das coordenadas está condicionada a da respectiva rede geodésica de apoio, que materializa o sistema de referência utilizado e serve de base para a determinação delas. Portanto, o desenvolvimento de técnicas que busquem maximizar a robustez de redes geodésicas é relevante. As investigações da presente pesquisa configuram contribuições para tal questão, no caso de redes altimétricas.

O Brasil, por intermédio do Instituto Brasileiro de Geografia e Estatística (IBGE), como um dos países participantes da IAG, visando contribuir com a materialização do IHRS, deverá adequar sua rede altimétrica às suas especificações, o que, obviamente, passa por atividades de ajustamento da mesma.

Para esse e outros processamentos de redes altimétricas, um procedimento de estimação da rede deve ser adotado. Com isso, espera-se que os resultados dessa pesquisa contribuam para a definição de metodologias para a implantação de redes altimétricas a nível nacional e internacional.

Por fim, nota-se na literatura que muitas vezes conceitos relacionados a diferentes etapas de um processo de estimação robusta de redes altimétricas são abordados de forma específica e pontual. Assim, para facilitar o entendimento do assunto de uma forma mais geral, foi possível verificar uma integração desses nessa pesquisa.

A descrição das contribuições dessa pesquisa no que se refere à estimação de redes geodésicas foi apresentada na seção 5.1.

1.4 ORGANIZAÇÃO DA DISSERTAÇÃO

Nessa dissertação, organizada em 7 capítulos, foram vistos os conceitos relativos à estimação robusta de redes altimétricas, com ênfase nas etapas relacionadas com as hipóteses

da pesquisa. Parte do conteúdo é referente a redes geodésicas em geral, e não somente às altimétricas. Assim, os conceitos específicos para redes altimétricas e os para geodésicas em geral foram devidamente identificados. Obviamente, tudo que foi discutido em relação a redes geodésicas em geral vale também para as altimétricas, sem que a recíproca seja necessariamente verdadeira.

O Capítulo 1 trata da introdução ao trabalho e apresenta uma contextualização sobre os problemas a serem tratados, as hipóteses, os objetivos, a justificativa e a organização do mesmo.

A fundamentação teórica da pesquisa é vista nos Capítulos 2, 3 e 4. No Capítulo 2 são abordadas a definição e a formulação da minimização de normais vetoriais de interesse, via programação linear, no ajustamento de observações. O Capítulo 3 é dedicado ao tratamento de *outliers*, uma das principais etapas da estimação robusta de redes geodésicas e contexto da aplicação da ML1. Já o Capítulo 4, após definir um conceito mais amplo para a estimação robusta de redes, não limitado ao tratamento de *outliers*, traz outros detalhes relevantes para tal estimação. Em especial, quanto à alternativa de modelo estocástico para o ajustamento pelo MMQ, a partir do resultado da ML_{∞} , a ser investigada nessa pesquisa.

Os procedimentos propostos nesta pesquisa, assim como os métodos de avaliação dos mesmos, são apresentados no Capítulo 5. O Capítulo 6 trata dos experimentos, resultados e análises referentes aos procedimentos propostos. Por fim, no Capítulo 7 constam as conclusões da pesquisa realizada e as sugestões para trabalhos futuros. O trabalho tem ainda três apêndices.

2 NORMAS L1 E L ∞ NO AJUSTAMENTO DAS OBSERVAÇÕES

Nesse Capítulo a formulação e as características das normas vetoriais L1 e L ∞ são apresentadas. As respectivas formulações para solução via programação linear, no ajustamento de observações, também são descritas.

2.1 CARACTERÍSTICAS DA APLICAÇÃO DAS NORMAS L1 E L ∞

Seja \mathbf{x} um vetor qualquer no espaço \mathbf{R}^m . Uma norma $\|\mathbf{x}\|$ desse vetor é uma função $\mathbf{f}:\mathbf{R}^m \rightarrow \mathbf{R}$ que satisfaça as propriedades das EQ 2.1 a 2.3. De interesse para essa pesquisa, as normas vetoriais L1, L2 (ou norma euclidiana), L ∞ (ou norma de *Chebyshev*) e de *Mahalanobis* são definidas pelas expressões das EQ 2.4 a 2.7, respectivamente. As EQ 2.1 a 2.6 foram adaptadas de GOLUB e VAN LOAN (1996), enquanto a EQ 2.7 o foi de XU *et al.* (2012). É fácil verificar que a norma L2 é um caso particular da norma de *Mahalanobis*, no qual \mathbf{D} é a matriz identidade de mesmas dimensões.

$$f(x) \geq 0, \quad x \in \mathbf{R}^m \quad (\text{EQ 2.1})$$

$$f(x + y) \leq f(x) + f(y), \quad x, y \in \mathbf{R}^m \quad (\text{EQ 2.2})$$

$$f(\alpha x) = |\alpha|f(x), \quad \alpha \in \mathbf{R}, x \in \mathbf{R}^m \quad (\text{EQ 2.3})$$

$$\|\mathbf{x}\|_1 = |x_1| + \dots + |x_m| \quad (\text{EQ 2.4})$$

$$\|\mathbf{x}\|_2 = (x_1^2 + \dots + x_m^2)^{1/2} = (\mathbf{x}^T \mathbf{x})^{1/2} \quad (\text{EQ 2.5})$$

$$\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq m} |x_i| \quad (\text{EQ 2.6})$$

$$\|\mathbf{x}\|_M = (\mathbf{x}^T \mathbf{D} \mathbf{x})^{1/2}, \quad \mathbf{D} \in \mathbf{R}^{m \times m} \text{ e } \mathbf{D} \text{ é positiva semi - definida} \quad (\text{EQ 2.7})$$

No consagrado ajustamento de observações pelo MMQ, para o caso de não se considerar os pesos das observações (ou seja, atribuir-se pesos unitários para todas elas), aplica-se a minimização da norma L2 ao vetor \mathbf{v} dos resíduos do ajustamento. Entretanto, as observações geodésicas em geral possuem diferentes precisões e, com isso, recebem diferentes pesos no ajustamento. Assim, para o caso geral de observações correlacionadas e/ou com pesos diferentes entre si, aplica-se a minimização da norma de *Mahalanobis*, também chamada de minimização da norma L2 ponderada (ou, até mesmo, de minimização da norma L2 somente), sobre o vetor \mathbf{v} , com a matriz dos pesos das observações \mathbf{P} no lugar da matriz \mathbf{D} realizando o ponderamento das mesmas. Mais detalhes sobre a matriz \mathbf{P} em (GHILANI, 2010).

Para aplicação das normas $L1$ e $L\infty$ no ajustamento de observações independentes ou decorrelacionadas (matriz \mathbf{P} diagonal), a minimização se dá no vetor \mathbf{pv} (EQ 2.8), construído pela multiplicação do módulo de cada resíduo \mathbf{v}_i pelo peso \mathbf{p}_i da respectiva observação.

$$pv = \begin{bmatrix} p(1) * |v(1)| \\ p(2) * |v(2)| \\ \vdots \\ p(m) * |v(m)| \end{bmatrix}, m \text{ é a quantidade de observações} \quad (\text{EQ 2.8})$$

A aplicação da norma $L1$ (ou $ML1$) corresponde à minimização da soma dos elementos de \mathbf{pv} no ajustamento. Já a aplicação da $L\infty$ (ou $ML\infty$) corresponde à minimização do máximo resíduo absoluto ponderado (MRAP) do ajustamento, ou seja, do maior elemento de \mathbf{pv} .

Sendo \mathbf{m} a quantidade de observações da amostra, $\mathbf{P}_{m \times m}$ a matriz de pesos das observações independentes e $\mathbf{p}_{m \times 1}$ o vetor com os elementos da diagonal principal de \mathbf{P} , as EQ 2.9 a 2.11 expressam, respectivamente, a $ML1$, o MMQ e a $ML\infty$ no ajustamento das observações. Reparar que na $ML\infty$ a multiplicação vetorial é elemento a elemento. Essas equações foram adaptadas de (AMIRI-SIMKOOEI, 2003), (GHILANI, 2010) e (MASON e HANDSCOMB, 2003), respectivamente.

$$ML1: \text{mín} (p^T |v|) \quad (\text{EQ 2.9})$$

$$MMQ: \text{mín} (v^T P v) \quad (\text{EQ 2.10})$$

$$ML\infty: \text{mín} \{ \max_{1 \leq i \leq m} (p_i .* |v_i|) \} \quad (\text{EQ 2.11})$$

Assim como para o MMQ , no ajustamento pela $ML1$ e pela $ML\infty$ também se pode não levar em conta pesos para as observações. Nesse caso, o vetor \mathbf{pv} , no qual a respectiva norma será aplicada, corresponde ao próprio vetor \mathbf{v} dos resíduos do ajustamento.

Para o caso de haver relação de dependência entre as observações da rede e, assim, \mathbf{P} não ser diagonal, uma solução para diagonalização e normalização de \mathbf{P} por decomposição de *Cholesky* para viabilizar o ajustamento pela $ML1$ foi apresentada por YETKIN e INAL (2011). Contudo, de interesse para o corrente trabalho, em redes altimétricas, é comum que as observações sejam independentes entre si, o que garante a ocorrência de matriz \mathbf{P} decorrelacionada (diagonal) e viabiliza a obtenção imediata do vetor \mathbf{p} . Assim, uma análise mais minuciosa de tal questão foge ao escopo dessa pesquisa.

Inspirado em ABDELMALEK e MALEK (2008), outras características da minimização das normas $L1$ e $L\infty$ (e da norma $L2$, para comparação) no ajustamento podem ser resumidas pela TAB. 2.1. As respectivas explicações a seguir também foram baseadas em KHARAKULEVA (2013).

A questão da maior resistência a *outliers* da norma L_1 foi indicada no Capítulo 1. Já as formulações para as soluções da ML_1 e da ML_∞ por programação linear serão abordadas na seção seguinte. Nessa pesquisa, os problemas de programação linear serão resolvidos pelo método *SIMPLEX*. Para um aprofundamento sobre programação linear e o método *SIMPLEX*, recomenda-se (DANTZIG, 1963). Um resumo sobre o tema é apresentado por (AMIRI-SIMKOOEI, 2003).

TAB. 2.1: Outras características da minimização das normas vetoriais de interesse

Norma Minimizada	Resiste a outliers?	Solução mais usual	Sempre única? “caso geral”	Repete resíduos ponderados?
L_1 (ML_1)	Sim	Por programação linear.	Não	Sim
L_2 (MMQ)	Não	Operações matriciais mais “simples”.	Sim	Não necessariamente
L_∞ (ML_∞)	Não	Por programação linear.	Não	Sim

O “caso geral” relacionado à unicidade de solução ótima da TAB 2.1 trata daqueles nos quais \mathbf{A} é de posto completo, sendo \mathbf{A} a matriz dos coeficientes das incógnitas de um sistema linear inconsistente, caso do ajustamento de redes de nivelamento. Vale ressaltar que se \mathbf{A} não possuir posto completo, nenhum dos ajustamentos (nem mesmo pelo MMQ) possuirá garantia de solução única. Com isso, a aplicação direta de nenhum deles asseguraria a ocorrência de um problema bem-posto (“*well-posed problem*”), caracterizado por possuir solução única e estável. Detalhes sobre problemas bem-postos em (WURTEMBERG, 2011).

Considerando a ocorrência do “caso geral”, a solução do MMQ é sempre única, porém o mesmo não vale para a ML_1 e a ML_∞ . Mesmo sendo de posto completo, se \mathbf{A} não atender à condição de *Haar* (para $\mathbf{A}_{m \times n}$ todo conjunto de n linhas de \mathbf{A} não é singular), os ajustamentos pela ML_1 e pela ML_∞ ainda são susceptíveis a não possuir solução ótima única.

Pela geometria das redes e por se valerem de observações redundantes, é razoável se esperar que a condição de *Haar* não seja satisfeita no caso de redes geodésicas, diminuindo a possibilidade de unicidade de solução para a ML_1 e a ML_∞ .

Sobre a repetição de resíduos ponderados, trata-se da propriedade da ML_1 e da ML_∞ de gerar resíduos ponderados de observações ajustadas com valores repetidos. No caso da ML_1 , parte dos resíduos apresenta exatamente o valor zero (e, logo, os respectivos resíduos

ponderados também são nulos), ou seja, a solução é esparsa. Já para a ML^∞ , a repetição ocorre com o MRAP do ajustamento (que é, também, o próprio valor mínimo da norma de L^∞).

Quando a solução ótima é única, tem-se que no mínimo k resíduos ponderados são nulos e no mínimo $k+1$ resíduos ponderados são iguais ao MRAP, respectivamente, para a $ML1$ e a ML^∞ , sendo k o posto da matriz A . Para o caso de múltiplas soluções ótimas, a solução via programação linear pelo método *SIMPLEX*, que será o caso dessa pesquisa, implica a ocorrência da respectiva quantidade mínima de resíduos ponderados nulos para a $ML1$. Já para a ML^∞ , o *SIMPLEX* não garante sua respectiva quantidade mínima apresentada de resíduos ponderados iguais ao MRAP.

2.2 FORMULAÇÃO EM REDES GEODÉSICAS

Nessa seção, a formulação do ajustamento de observações pela $ML1$ e pela ML^∞ , via programação linear para implementação pelo método *SIMPLEX*, será apresentada, após breve revisão do MMQ. Em todos os casos, será considerado que o sistema de equações é linear e que as observações a serem ajustadas são independentes (a matriz dos pesos P é diagonal). A ocorrência de tais premissas é algo comum em redes altimétricas e será o caso nos experimentos dessa pesquisa.

O modelo funcional paramétrico do ajustamento de redes geodésicas pode ser definido pela EQ 2.12, sendo $A_{m \times n}$ a matriz dos coeficientes do vetor $x_{n \times 1}$ dos parâmetros da rede (incógnitas), $l_{m \times 1}$ o vetor dos valores observados, $v_{m \times 1}$ o dos resíduos, m a quantidade de observações e n a de parâmetros. Usualmente, no modelo estocástico, arbitra-se que o fator de variância *a priori* σ_0^2 é unitário e, assim, a matriz dos pesos das observações $P_{m \times m}$ é dada pela inversa da matriz variância-covariância (MVC) das mesmas $\Sigma L b_{m \times m}$ (EQ 2.13). A aplicação do MMQ ao sistema consiste em minimizar a soma dos quadrados dos resíduos ponderados pela matriz P (EQ 2.10) e resulta na solução para o vetor x (EQ 2.14). Maiores detalhes sobre modelo funcional e estocástico em (GHILANI, 2010) e sobre fator de variância *a priori* e MVCs no ajustamento pelo MMQ em (GEMAEL *et al.*, 2015).

$$Ax = l + v \quad (EQ\ 2.12)$$

$$P = \sigma_0^2 * (\Sigma L b)^{-1} = (\Sigma L b)^{-1}, \text{ com } \sigma_0^2 = 1 \text{ (adimensional)} \quad (EQ\ 2.13)$$

$$MMQ: x = (A^T P A)^{-1} A^T P l \quad (EQ\ 2.14)$$

2.2.1 FORMULAÇÃO DA ML1

Para o ajustamento de redes geodésicas pela ML1 (EQ 2.9), a soma dos módulos dos resíduos ponderados pelos respectivos pesos de suas observações independentes, por meio do vetor de pesos \mathbf{p}_{mx1} (composto pelos elementos da diagonal principal de \mathbf{P}), deve ser mínima. Para solução desse problema de minimização por programação linear, é necessário que todas variáveis \mathbf{x}_i e \mathbf{v}_i (componentes dos vetores \mathbf{x} e \mathbf{v} respectivamente) sejam não negativas. Para contornar tal limitação, dado que resíduos e parâmetros da rede, na prática, não possuem restrição de sinal, devem-se inserir variáveis artificiais para \mathbf{x} e \mathbf{v} (EQ 2.15). Assim, após se resolver o problema para as variáveis não negativas α_i , β_i , u_i e w_i e retornar-se às originais \mathbf{x}_i e \mathbf{v}_i , essas últimas poderão assumir valores negativos. Os vetores α e β têm as mesmas dimensões de \mathbf{x} , e u e w as de \mathbf{v} . Contudo, a função objetivo Z_{ML1} a ser minimizada passa a ser a da EQ 2.16.

$$x = \alpha - \beta; v = u - w \quad (\text{EQ 2.15})$$

$$Z_{ML1} = \text{mín} (p^T |u - w|) \quad (\text{EQ 2.16})$$

A EQ 2.17, adaptada de (AMIRI-SIMKOOEI, 2003), representa as restrições do problema de programação linear, sendo \mathbf{I}_{mxm} uma matriz identidade. Ela equivale ao próprio modelo funcional do ajustamento de redes geodésicas da EQ 2.12, com as adaptações para resolução pelo *SIMPLEX*.

$$\begin{bmatrix} A & -A & -I & I \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ u \\ w \end{bmatrix} = [l] \quad (\text{EQ 2.17})$$

Entretanto, voltando à função objetivo, para um determinado resíduo \mathbf{v}_i , além das respectivas variáveis artificiais u_i e w_i serem não negativas, pode-se garantir que no mínimo uma delas será nula, pois ambas jamais farão parte de uma solução básica simultaneamente pelo método *SIMPLEX*. Isso, pois, nas restrições do problema, suas colunas serão sempre linearmente dependentes, por serem correspondentes em \mathbf{I}_{mxm} e $-\mathbf{I}_{mxm}$ (WATSON, 2000). A inclusão das colunas de zeros abaixo dessas matrizes (algo específico para o caso de injunções em redes geodésicas, não considerado na obra citada) na EQ 2.17 não altera a condição de dependência linear.

Portanto, a função objetivo Z_{ML1} a ser minimizada pode ser expressa pela EQ 2.18, que, no formato padrão da programação linear, é dada pela EQ 2.19, sendo $\mathbf{0}_{mx1}$ um vetor de zeros.

Assim, nota-se que a função objetivo tem como incógnitas α e β com coeficiente nulo e u e w com coeficientes dados por p .

$$Z_{ML1} = \min \{p^T(u + w)\} \quad (\text{EQ 2.18})$$

$$Z_{ML1} = [0^T \quad 0^T \quad p^T \quad p^T] \begin{bmatrix} \alpha \\ \beta \\ u \\ w \end{bmatrix} \rightarrow \min \quad (\text{EQ 2.19})$$

2.2.2 FORMULAÇÃO DA ML_∞

Já para o ajustamento de redes geodésicas pela ML_∞ , o maior valor absoluto dentre os resíduos ponderados deve ser minimizado (EQ 2.11). Para solução via programação linear, deve-se inserir variáveis artificiais para x e v (EQ 2.15), assim como feito na $ML1$, visando permitir valores negativos para seus elementos. Após, cria-se uma variável unidimensional s , a qual deve ser minimizada (ABDELMALEK e MALEK, 2008) e caracteriza a função objetivo Z_{ML_∞} (EQ 2.20), sendo $\mathbf{0}_{2(n+m) \times 1}$ um vetor de zeros e $[\mathbf{1}]_{1 \times 1}$ uma matriz cujo único elemento é “1”.

$$Z_{ML_\infty} = [0^T \quad [1]] \begin{bmatrix} \alpha \\ \beta \\ u \\ w \\ s \end{bmatrix} \rightarrow \min \quad (\text{EQ 2.20})$$

A estratégia é impor que o módulo de cada um dos resíduos v_i (substituídos por $u_i - w_i$) multiplicados pelos pesos das respectivas observações p_i (dados pelos elementos $P_{i,i}$ de P) sejam menores ou iguais a s , sendo, assim, restrições do problema de programação linear (EQ 2.21). Nessa equação matricial, $\mathbf{0}_{m \times n}$ (lado esquerdo da inequação) são matrizes de zeros, $\mathbf{0}_{2m \times 1}$ (lado direito da inequação) é um vetor de zeros, $P_{m \times m}$ a matriz dos pesos das observações e $[-1]_{m \times 1}$ um vetor cujos elementos são todos iguais a “-1”.

$$\begin{bmatrix} 0 & 0 & -P & P & [-1] \\ 0 & 0 & P & -P & [-1] \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ u \\ w \\ s \end{bmatrix} \leq 0 \quad (\text{EQ 2.21})$$

Por fim, a EQ 2.17, apresentada na explicação da $ML1$, também será aplicada como restrição na ML_∞ . Para que seja consistente com as demais restrições, no caso da ML_∞ , foi

necessário incluir uma coluna $\mathbf{0}_{s(\mathbf{m} \times 1)}$ para a variável \mathbf{s} , mas com todos os elementos iguais a zero, pois a mesma não participa do modelo funcional (EQ 2.22).

$$\begin{bmatrix} A & -A & -I & I & \mathbf{0}_s \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ u \\ w \\ s \end{bmatrix} = [l] \quad (\text{EQ 2.22})$$

Vale ressaltar que, tanto para a ML1 como para a ML^∞ , apesar dos parâmetros da rede \mathbf{x}_i (substituídos por α_i e β_i) possuírem coeficientes nulos na função objetivo (que representa aquilo que se deseja minimizar), eles não deixam de ser incógnitas (uma vez que se adotou o modelo funcional paramétrico), sendo, portanto, estimados na solução dos respectivos problemas de programação linear.

3 CONTEXTO DA APLICAÇÃO DA NORMA L1

Devido à sensibilidade do ajustamento pelo MMQ aos *outliers*, torna-se necessário dar-lhes algum tipo de tratamento, de forma a diminuir ou até isolar seus efeitos nos parâmetros a serem estimados. Logo, o tratamento de *outliers* é essencial para um apropriado ajustamento de observações geodésicas. Nesse sentido, pelos motivos já expostos (ver TAB 2.1), nessa pesquisa a aplicação do ajustamento pela ML1 refere-se à identificação de *outliers* no conjunto de observações.

É improvável que se possa garantir a ausência de *outliers* em dados coletados em campo na prática dos levantamentos (KNIGHT *et al.*, 2010). Ademais, segundo LEHMANN (2013b), o custo de se evitar totalmente erros grosseiros (mais frequentes causadores de *outliers* em redes geodésicas) é economicamente injustificável, sendo mais viável aceitar o risco da existência de alguns para posterior tratamento. Assim, o tratamento de *outliers*, que será visto com maiores detalhes na seção 3.2, sempre deve ser realizado.

Neste contexto, GEMAEL *et al.* (2015) alertam que erros grosseiros podem acontecer até mesmo no processamento eletrônico de equipamentos, como nos rastreadores de posicionamento por satélite e níveis eletrônicos. BUSTOS (1981) conclui que a existência de uma pequena quantidade de *outliers* em observações, considerando diversas aplicações, incluindo a engenharia, parece ser a regra, e não a exceção.

Entretanto, visando o tratamento de *outliers* propriamente dito, uma análise conceitual dos termos relacionados, seus significados e aspectos é necessária para que se tenha um entendimento consistente e homogêneo sobre os mesmos. A seguir, algumas questões sobre o tema foram levantadas, especificamente quanto aos erros das observações gerando *outliers* em redes geodésicas. Assim, para tratar somente dessa causa específica, parte-se da premissa de que, dentre outros, não há problemas na modelagem matemática do ajustamento.

3.1 OUTLIERS X ERROS GROSSEIROS

Os erros em observações geodésicas são classificados em aleatórios, sistemáticos e grosseiros. Os aleatórios (também chamados de acidentais ou randômicos) são os inevitáveis erros de medida, únicos presentes em todas as observações geodésicas. Sua distribuição de frequência tende a se aproximar da distribuição normal, com aumento do tamanho da amostra. Os sistemáticos tratam de condições da coleta das observações que podem ser modeladas e ter

seus efeitos neutralizados ou minimizados, ora com procedimentos de coleta, ora com modelos matemáticos apropriados.

Já os erros grosseiros (também conhecidos por *blunders*, *mistakes*, *gross errors* ou *bad errors*) são os que não são aleatórios, nem sistemáticos. Não possuem distribuição nem modelagem conhecida e podem ser de qualquer magnitude. Em geral ocorrem pontualmente, por falhas humanas ou de equipamento. Maiores detalhes sobre erro grosseiro, sistemático e aleatório em (GEMAEL *et al.*, 2015) e (GHILANI, 2010).

Nessa pesquisa, o termo “erro total” e_T (EQ 3.1) será utilizado para denotar a soma dos erros aleatório e_A (que sempre existe), sistemático e_S (caso exista) e grosseiro e_G (caso exista) de uma observação.

$$e_T = e_A + e_S + e_G \quad (\text{EQ 3.1})$$

As observações geodésicas o_G (EQ 3.2) são compostas pelo seu valor exato v_E (desconhecido na prática dos levantamentos) da grandeza medida e seu erro total. Podem ser classificadas em “boas” observações (“*good*”, *useful observations* ou *inliers*) ou *outliers*. Assim, as primeiras são todas que não são *outliers*.

$$o_G = v_E + e_T \quad (\text{EQ 3.2})$$

Uma possível tradução para o português de *outlier* é a de “valor atípico”, porém o termo tem sido mantido, sem tradução, em trabalhos na língua portuguesa. Segundo KLEIN (2014), a definição mais citada na literatura é a de HAWKINS (1980), pela qual “*an outlier is an observation that deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism*”.

Da definição apresentada, se nota que *outlier* não é um tipo de erro (é uma observação), nem está relacionado a algum tipo de erro específico. Assim, erros grosseiros, sistemáticos (não modelados) e até aleatórios (ou combinações deles) podem, em teoria, ser a causa de *outliers*, desde que o desvio da observação faça com que a mesma “levante a suspeita de um mecanismo diferente”. Logo, *outlier* e erro grosseiro são conceitos distintos.

Lembrando que erros aleatórios estão em todas as observações geodésicas, a FIG 3.1 ilustra o posicionamento dos *outliers* dentre elas. Enfim, todas as combinações de erros podem gerar *outliers* ou “boas” observações, pois o critério de classificação está na discrepância em relação à amostra, não nos tipos de erro que a observação contém.

Apesar disso, os conceitos de *outlier* e erro grosseiro são por vezes confundidos. Segundo LEHMANN (2013a), isso ocorre pois “*in geodesy, outliers are most often caused by gross errors and gross errors most often cause outliers*”. Ou seja, apesar das ocorrências mais comuns contribuírem para esse equívoco conceitual, há exceções a esse caso geral. Logo, nem

todo *outlier* é causado por erro grosseiro e nem todo erro grosseiro causa *outlier*. Ainda nessa referência, *outlier* é definido como “*an observation that is so probably caused by a gross error that it is better not used or not used as it is*”. Vale notar que, também por ela, um *outlier* pode ou não conter erro grosseiro, pois o critério é a probabilidade disso ter ocorrido, e não a ocorrência de fato.

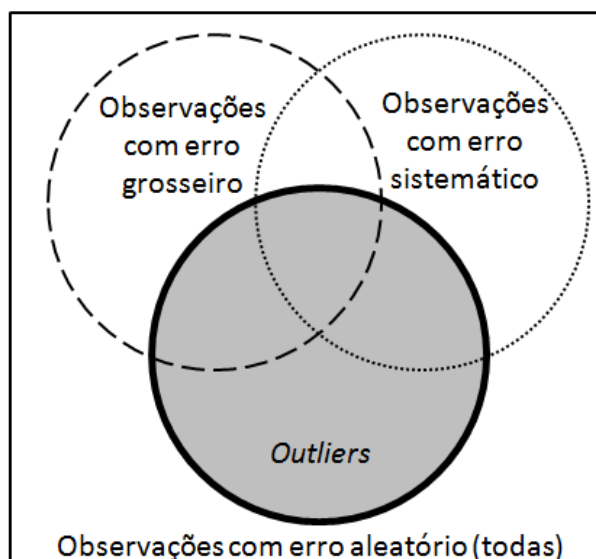


FIG. 3.1: Posicionamento de *outliers* dentre as observações geodésicas.

Assim sendo, tanto HAWKINS (1980) quanto LEHMANN (2013a) parecem concordar no que se refere ao seguinte aspecto: a grande probabilidade de erro grosseiro pode ser vista como o fato que levanta a suspeita de um mecanismo diferente. Na realidade, fala-se em probabilidade de erro grosseiro (e não na ocorrência propriamente), até porque não é possível se determinar, desprovida de incerteza, a ocorrência de erro grosseiro na prática dos levantamentos. BAARDA (1968), principal referência relacionada às aplicações geodésicas, reconhece que “*because of the random character of observations it is impossible to signalize gross erros with certainty*”.

Sob o risco de ser repetitivo, é fundamental para encadeamento das ideias que se perceba a sutil - mas relevante - diferença entre a observação “ter grande probabilidade de conter erro grosseiro” (levantar a suspeita do mecanismo diferente) e “conter um erro grosseiro”. Além disso, é razoável se imaginar que observações com erros totais relativamente maiores tendam a apresentar maior probabilidade de conter erro grosseiro, mesmo que não contenham.

Assim, por exemplo, uma observação sem erro grosseiro e sem erro sistemático pode ter significativa discrepância devido a um grande desvio em seu erro aleatório – e essa ocorrência

estar gerando *outlier*. Apesar de pouco provável, isso não é impossível. Deve-se lembrar de que os erros aleatórios das observações tendem a possuir distribuição normal, cuja função densidade, apesar de relativamente maior nas regiões mais próximas da abscissa zero, é definida e não nula em todo o espaço \mathbf{R} dos números reais.

Além disso, uma observação sem efeitos sistemáticos e com erros aleatório e grosseiro de magnitude relativamente “pequena” provavelmente não será discrepante, não configurando um *outlier*, apesar de conter erro grosseiro.

Contudo, conclui-se que erro grosseiro e *outlier* não são sinônimos, nem tampouco um é caso particular do outro, pois nem sempre um erro grosseiro causa um *outlier* e nem todo *outlier* é causado por erro grosseiro.

3.1.1 CONSIDERAÇÕES SOBRE *OUTLIERS* DE FORMA OBJETIVA

As definições de *outlier* apresentadas são mais teóricas do que práticas. De forma mais prática, para o caso unidimensional, uma das opções adotada nas ciências em geral é tomar-se como *outliers* as observações que desviam mais de determinada quantidade de vezes do seu desvio padrão σ em relação à própria média μ . Isso é baseado nas características da distribuição normal, que apresenta 99,73% e 95,44% dos dados, para os intervalos $\mu \pm 3\sigma$ e $\mu \pm 2\sigma$, dois dos mais usuais, respectivamente. A adoção do primeiro intervalo para identificação de *outliers* é usualmente chamada de *regra dos 3 σ* . Segundo LEYS *et al.* (2013), nesse contexto, apesar de haver autores que defendem diferentes quantidades de σ para o limite do desvio, essa escolha depende da situação e da perspectiva defendida pelo pesquisador.

Contudo, nota-se que, também nessa abordagem objetiva para o termo *outlier*, corroborando a conclusão da seção anterior, a classificação de uma observação como *outlier* não depende do tipo de erro que a mesma contém, mas da magnitude e da tolerância aceita do desvio de seu erro total em relação ao das demais.

Na estimação de redes geodésicas, já tratando de casos multidimensionais, o trabalho de BAARDA (1968) foi pioneiro na colocação de uma abordagem objetiva para o tema. Desenvolveu o *data-snooping* (seção 3.2.1), um teste estatístico de hipótese para identificação de erros grosseiros em observações. Após tratamento de seus efeitos sistemáticos, observações rejeitadas pelo teste passaram a ser classificadas como *outliers*. Assim, por exemplo, uma observação sem erro grosseiro, mas rejeitada no teste, também é considerada *outlier*.

Nos últimos 50 anos, por meio de métodos alternativos para identificação de *outliers*, outras abordagens objetivas para o termo vieram na literatura geodésica, sejam oriundos de outros testes estatísticos, ou de técnicas de estimação robusta (seção 3.2). Para se avaliar a eficácia desses métodos alternativos, *outliers* passaram a ser gerados nas redes geodésicas pela inserção proposital de erros grosseiros com “grande” magnitude, suficiente para tornar a respectiva observação discrepante. Assim, o método obtém êxito caso identifique corretamente esse suposto *outlier*. Lembrando a questão de que a maioria dos *outliers* em Geodésia são causados por erros grosseiros, apesar desse procedimento não cobrir todos os casos possíveis de *outlier*, corresponde ao mais comum.

Nesse contexto, diversos autores consideram o mínimo de $3\sigma_i$, sendo σ_i o respectivo desvio padrão da observação, para que o erro grosseiro passe a ser considerado “grande” o suficiente para causar um *outlier*. Essa abordagem será chamada nessa pesquisa de *regra dos 3σ para o erro grosseiro*. Exemplos da inserção de *outliers* pela *regra dos 3σ para o erro grosseiro* para avaliação de métodos de identificação de *outliers* podem ser encontrados em (HEKIMOGLU e ERENOGLU, 2007), (KLEIN *et al.*, 2015) e (DURDAG *et al.*, 2016). Ela é similar à *regra dos 3σ* apresentada anteriormente, porém aplicando o limite de 3σ a um possível erro grosseiro da observação, não ao erro total da mesma.

Pelo conteúdo apresentado, todos os tipos de erro e combinações deles podem gerar *outliers*. Porém, ao se adotar a *regra dos 3σ para o erro grosseiro*, se coloca que *outliers* são causados por erros grosseiros maiores que 3σ (em módulo). Com isso, se desprezam, por exemplo, duas situações que parecem merecer maior atenção: 1ª) que o erro aleatório pode ser discrepante a ponto de gerar *outlier*; 2ª) que uma observação pode não ter grande probabilidade de conter erro grosseiro, não sendo considerada *outlier*, mesmo que contenha erro grosseiro com valor absoluto maior que 3σ .

Quanto ao primeiro ponto levantado, a área sob a curva normal padrão limitada por $\pm 3\sigma$ é de 0,9973. Com isso, mesmo que não haja erros grosseiros ou sistemáticos, há probabilidade de 0,27% do erro total de uma observação isoladamente extrapolar os $\pm 3\sigma$, devido ao seu erro aleatório. Nesse caso a observação possui erro discrepante (FIG 3.2), o que sugere que a mesma deva ser classificada como *outlier* (o que não seria o caso pela *regra dos 3σ para o erro grosseiro*), ainda que na realidade (desconhecida na prática) tenha somente erro aleatório.

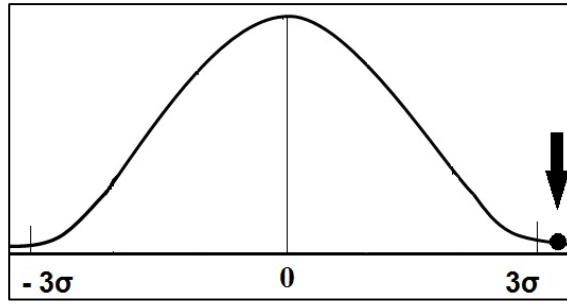


FIG. 3.2: Erro aleatório discrepante.

Vale ressaltar que, apesar da probabilidade de 0,27% para uma observação isoladamente, em cenários com quantidade relativamente alta de observações, como a Rede Altimétrica de Alta Precisão (RAAP) do Sistema Geodésico Brasileiro, a probabilidade **prob** do erro aleatório de pelo menos uma observação ter valor absoluto maior que 3σ aumenta significativamente. Para redes com 1.000 observações, por exemplo, tal probabilidade já é de 93,30% (EQ 3.3).

$$prob_{1.000} = 1 - (99,73/100)^{1.000} = 93,30\% \quad (\text{EQ 3.3})$$

Já quanto à segunda situação levantada no antepenúltimo parágrafo, por exemplo, uma observação com erro grosseiro de $3,1\sigma$ (acima dos 3σ , ou seja, que seria classificada como *outlier* pela *regra dos 3σ para o erro grosseiro*), mas com erro aleatório de $2,0\sigma$ de sinal contrário, tem erro total ($1,1\sigma$) não discrepante (FIG 3.3), pelo limite considerado de 3σ . Assim, espera-se que “não levante a suspeita de mecanismo diferente” e “não possua grande probabilidade de erro grosseiro”, não devendo, portanto, ser considerada *outlier*.

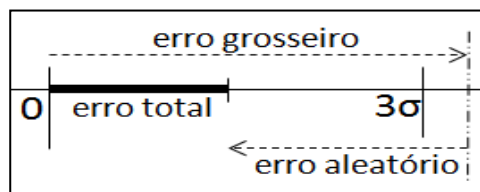


FIG. 3.3: Erro total não discrepante, apesar de erro grosseiro acima do limite.

Assim, é razoável se entender que a *regra dos 3σ* não deva valer apenas para o erro grosseiro das observações, mas sim para o erro total das mesmas, independentemente dos tipos de erro que ela contém. Ou seja, um *outlier* poderia ser definido como “uma observação com erro total de magnitude superior a $3\sigma_i$, sendo σ_i o seu desvio padrão”. Esse entendimento será chamado nessa pesquisa de *regra dos 3σ para o erro total*.

Com isso, observações com erros aleatórios discrepantes em módulo (acima dos 3σ) seriam consideradas *outliers* e aquelas com erro total dentro dos limites dos 3σ , mesmo que eivadas de erro grosseiro, não seriam. Ou seja, possivelmente os dois problemas levantados anteriormente seriam contornados.

É imprescindível ressaltar ainda que, apesar do 3σ ser um limite usual, outros já foram ou podem ser testados. Na realidade, não há na literatura uma justificativa matematicamente rigorosa que coloque algum limite como ideal. Na escolha de limites objetivos para *outliers*, em consonância com o entendimento visto anteriormente de LEYS *et al.* (2013), a opção de ser mais estrito (que classifica uma observação como *outlier* quando a evidência para tal é mais forte) ou brando, respectivamente com aumento ou diminuição do intervalo de aceitação de observações, fica a cargo da avaliação do analista.

Contudo, a *regra dos 3σ para o erro total* foi adotada na inserção de erros para avaliação de métodos de identificação de *outliers* nessa pesquisa.

3.2 TRATAMENTO DE *OUTLIERS* EM REDES GEODÉSICAS

No tratamento de *outliers*, TEUNISSEN (2000) apresenta 3 fases: a) detecção, que corresponde à verificação genérica da existência de *outliers* na rede, sem identificação dos mesmos, dentre outros problemas no ajustamento; b) identificação, que é a indicação de quais observações são *outliers*; c) e, adaptação, que trata da ação corretiva com relação a *outliers* identificados. Inspirado nessa referência, a FIG 3.4 apresenta uma taxonomia para as fases do tratamento de *outliers* em redes geodésicas.

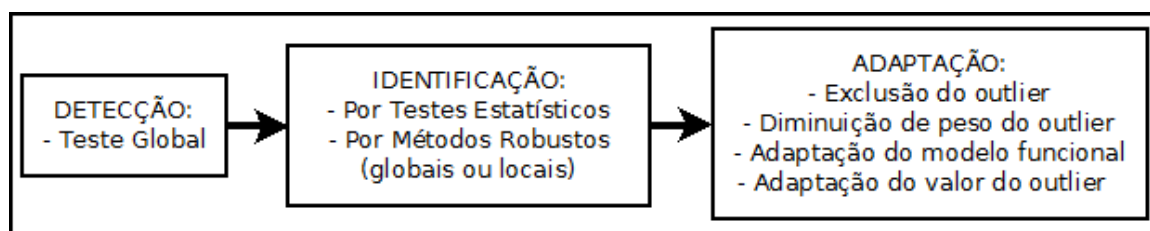


FIG. 3.4: Fases do tratamento de *outliers*.

Entretanto, é fundamental ressaltar que, para a identificação de *outliers* a partir de resultados do ajustamento pela ML1, por exemplo, dentre outras, não há na literatura a fase de detecção. Ou seja, parte-se direto para a identificação de *outliers*. Assim, a fase de detecção tem relação direta com a metodologia de tratamento de *outliers* daquela referência, não sendo

sua aplicação obrigatória, nem usual, em outros cenários de tratamento de *outliers*.

O Teste Global do Ajustamento, ou Teste *Qui-Quadrado*, é aplicado para verificação, além de outros problemas no ajustamento, da possível ocorrência de *outliers* na amostra de observações da rede. Uma de suas limitações é que ele pode apenas detectar a ocorrência de *outliers*, sem identificar precisamente em qual observação. Para maiores detalhes sobre o Teste *Qui-Quadrado*, que corresponde à fase de detecção na abordagem de TEUNISSEN (2000), sugere-se adicionalmente KLEIN (2011).

Para a identificação de *outliers* em redes geodésicas, há 2 principais tipos de abordagem: testes estatísticos e métodos robustos (HEKIMOGLU *et al*, 2011).

Os testes estatísticos são realizados a partir dos resultados de um ajustamento inicial da rede pelo MMQ. O DS (BAARDA, 1968), que corresponde à fase de identificação de TEUNISSEN (2000), é o mais usual deles para identificação de *outliers* em redes geodésicas. Pontos importantes para sua aplicação em redes altimétricas serão abordados na seção 3.2.1, pois seu desempenho será utilizado nessa pesquisa para comparação na avaliação da identificação de *outliers* a partir de resultados da ML1 que será investigada. Quanto aos testes estatísticos, podem ser citados ainda o Teste *Tau* (POPE, 1976) e o generalizado para múltiplos *outliers* (TEUNISSEN, 2000).

Já os métodos robustos utilizam estimadores mais resistentes a *outliers*. Assim, quando da sua aplicação em redes geodésicas, espera-se que os *outliers* apresentem resíduos de maior magnitude em relação às demais “boas” observações. Tal efeito foi visto, para o caso da ML1, na FIG 1.1. Em consequência, observações com resíduos considerados “grandes”, segundo algum critério especificado, podem ser identificadas como *outliers*.

Os métodos robustos, quando solucionados por técnicas de otimização global, podem ser chamados de globais. Por exemplo, o própria ML1, com solução por programação linear (que é uma técnica de otimização global), é considerado um método robusto global. Por ser uma das investigações desta pesquisa, a identificação de *outliers* pela ML1 será vista mais detalhadamente na seção 3.2.2. Outros estimadores globais já foram aplicados às redes geodésicas. Como exemplo, em (YETKIN e BERBER, 2013), o MMQ com restrição de sinal, no qual a soma dos sinais dos resíduos deve ser nula, por técnica de otimização global, foi aplicado a uma rede GNSS. Já em (YETKIN e BERBER, 2014), os métodos da mínima variância dos resíduos quadrados e dos mínimos quadrados “aparado”, que é similar ao MMQ, mas com a exclusão dos maiores resíduos ponderados, foram verificados.

Há, ainda, métodos robustos locais, os quais normalmente aplicam alguma estratégia de pesos iterativos (EPI), no qual a rede é sucessivamente ajustada pelo MMQ, sendo que, após

cada iteração, os pesos das observações são reponderados em função dos resíduos do ajuste anterior (a observações com maiores resíduos são atribuídos menores pesos), até que se chegue a certo nível de convergência para os resultados do ajustamento. Diferentes métodos possuem diferentes funções para atribuição dos pesos às observações. Assim, partindo da premissa de que, no ajustamento iterativo pelo MMQ, *outliers* tendem a ter resíduos relativamente maiores (em relação às “boas” observações, que não configuram *outliers*), ao final das iterações espera-se que seus pesos sejam relativamente menores (mas ainda significativos) ou até não significativos, respectivamente diminuindo ou isolando seus efeitos nos parâmetros estimados. São chamados locais, pois, se a solução ótima global do estimador robusto estiver defasada significativamente em relação à solução inicial pelo MMQ, poderá o resultado do estimador ser um ótimo apenas local, com qualidade relativamente inferior (YETKIN e BERBER, 2013). GOKALP *et al.* (2008), por exemplo, apresentam a formulação e a aplicação em rede GNSS de diversos estimadores robustos locais, via EPI, como os de *Danish*, de *Huber*, de *Andrews*, de *Turkey* e de *Yang-I*. TRASÁK e STRONER (2014) mostram comparações de desempenho dos estimadores locais de *Huber*, de *Hampel* e de *Andrews*. Vale citar ainda que comparações sobre realizações locais e globais de métodos robustos foram feitas por BASELGA (2007).

Por fim, a adaptação (ou ação corretiva), terceira – mas pode ser a segunda, função da abordagem realizada - fase do tratamento de *outliers* em redes geodésicas, corresponde à providência tomada em relação aos *outliers* identificados. Pode-se elencar 4 possibilidades de tratamentos, a saber:

- 1) A primeira opção é excluir o *outlier* da amostra, que é a ação mais usual após identificação pelo DS (TEUNISSEN, 2000);

- 2) outra opção é diminuir o seu peso, mas sem torná-lo insignificante no ajustamento (o que, obviamente, diminui sua influência nos parâmetros estimados da rede). Isto pode ser resultado, por exemplo, do reponderamento obtido por algum estimador robusto local, como ocorrido em (GOKALP *et al.*, 2008);

- 3) a terceira alternativa é adaptar o modelo funcional da rede pela inserção de parâmetros auxiliares, correspondentes aos erros grosseiros dos *outliers* identificados, como apresentado por (TEUNISSEN, 2000). Na prática, isso conduz aos mesmos resultados da exclusão do outlier, pois a observação perde sua redundância e não mais influencia as demais no ajustamento;

- 4) uma última opção que se poderia imaginar ainda seria a adaptação do valor da observação identificada como *outlier*. Isto dependeria de se saber o valor do erro grosseiro

dessa, algo pouco provável na prática geodésica, pois a separação (exata e desprovida de incerteza) entre o valor dela com respectivo erro aleatório (que sempre existe) e o do possível erro grosseiro é inviável, não havendo referências na literatura sobre tal. Entretanto, um exemplo de possível visualização dessa adaptação do valor seria a ocorrência de um erro grosseiro no sinal no valor de uma observação, passando, assim, a mesma a configurar um *outlier*. Com uma análise mais detalhada dessa observação em relação às suas vizinhas, pode ser que seja possível perceber que uma simples troca de sinal a tornaria consistente com seu valor esperado, descaracterizando a probabilidade da existência do *outlier*. Assim sendo, poder-se-ia assumir esse novo valor para a observação, apesar da incerteza envolvida na operação.

Vale ressaltar ainda que, em todos os casos de adaptação, o analista pode optar pela necessidade de coleta de novas observações em campo. Principalmente na primeira e terceira alternativas apresentadas, pois a exclusão de observações e a inserção de parâmetros diminuem a redundância da rede, essencial até mesmo para identificação de outros possíveis *outliers*.

3.2.1 O DATA SNOOPING

O trabalho de BAARDA (1968) foi pioneiro na identificação de *outliers* e na aferição da confiabilidade de redes geodésicas. Propôs o procedimento *data-snooping*, um teste estatístico baseado na razão de verossimilhança para identificação de observações suspeitas de serem *outliers*, por possuírem alta probabilidade de estarem eivadas de erro grosseiro. Ainda hoje é considerado um dos melhores métodos de identificação de *outliers* em redes geodésicas (ROFATTO *et al.*, 2017), sendo, ainda, o mais empregado em *softwares* comerciais e recomendado em livros-texto relacionados (LEHMANN e LOSLER, 2016), o que justifica sua utilização nessa pesquisa para comparação com o classificador que será proposto.

Para o caso da MVC das observações ser diagonal, algo comum em redes altimétricas e que será o caso dessa pesquisa, o cálculo da estatística do teste \mathbf{W}_i se baseia na EQ 3.4, sendo \mathbf{v}_i o resíduo da *i-ésima* observação e $\sigma_{\mathbf{v}_i}$ o desvio padrão (DP) desse no ajustamento pelo MMQ (TEUNISSEN, 2000). Os valores \mathbf{W}_i são comparados com os tabelados $-\mathbf{Z}_{\alpha/2}$ e $\mathbf{Z}_{\alpha/2}$ para a distribuição normal padrão (teste bicaudal, no qual α é o nível de significância) e utilizados para testar a hipótese nula de que a *i-ésima* observação não configura um *outlier*. Caso a condição da EQ 3.5 não seja satisfeita, a hipótese nula é rejeitada e a observação é considerada suspeita.

$$W_i = \frac{v_i}{\sigma_{v_i}} \quad (\text{EQ 3.4})$$

$$-Z_{\alpha/2} \geq W_i \geq Z_{\alpha/2} \quad (\text{EQ 3.5})$$

Como o método considera a existência de apenas 1 *outlier* por vez, após todas as observações serem testadas, a suspeita de maior W_i é retirada (normalmente a exclusão é a ação corretiva empregada) e a próxima iteração inicia, com um novo ajustamento pelo MMQ e aplicação do DS, até que não haja nenhuma observação suspeita (TEUNISSEN, 2000). Essa extensão do DS é chamada de DS iterativo. Por ser a mais usual, será chamada simplesmente de DS na sequência do presente trabalho. O nível de significância recomendado é de $\alpha=0,001$ para o DS (BAARDA, 1968). Esse mesmo valor foi considerado ótimo por ROFATTO *et al.* (2017), com base nos resultados de seus experimentos.

Relacionados ao DS estão ainda a teoria da confiabilidade, que trata das confiabilidades interna e externa da rede (BAARDA, 1968), e a análise de robustez (BERBER, 2006). Como não são aplicáveis à identificação de *outliers* pela ML1, e, assim, não poderiam ser comparadas, tais questões não foram alvos dessa pesquisa.

3.2.2 IDENTIFICAÇÃO DE *OUTLIERS* PELA ML1 E SEU ESTADO DA ARTE

Devido à já explorada deficiência de robustez do MMQ, pode ocorrer do ajustamento inicial por ele ter “suavizado” *outliers*, dificultando a identificação desses por testes estatísticos. Ademais, tais testes reconhecem *outliers* até determinado valor mínimo e com certo poder de teste associado. Portanto, podem ocorrer *outliers* não detectados pelos mesmos (HUBER e RONCHETTI, 2009).

Principalmente no caso de múltiplos *outliers*, sua influência no ajuste pelo MMQ pode acarretar, ainda, a não identificação de um deles (mascaramento, efeito *masking*), ou a errada identificação de uma observação “boa” (efeito *swamping*) (GUI *et al.*, 2011). Além disso, BASELGA (2011) indica a necessidade de hipóteses adicionais para que a determinação de *outliers* não seja ambígua, dado que diferentes combinações deles podem conduzir aos mesmos resíduos no ajuste pelo MMQ.

Conforme visto, até mesmo no caso da identificação de *outliers* por métodos robustos locais (via EPI), a rede é sucessivamente ajustada pelo MMQ. Ou seja, apesar de serem definidos como robustos, dependem também de um ajuste inicial da rede pelo MMQ (assim como os testes estatísticos), o qual pode ser irremediavelmente contaminado por *outliers*. Por conseguinte, surge o interesse pela investigação de métodos de identificação de *outliers* que

não sejam conduzidos a partir dos resíduos do MMQ. Neste contexto, a robustez da ML1 global (resolvido por métodos de otimização global) pode ser aproveitada.

Especificamente em redes geodésicas, já foi aplicado, por exemplo, por AMIRI-SIMKOOEI (2003) e YETKIN e INAL (2011). Em ambos os casos, a solução foi obtida pelo método *SIMPLEx* de programação linear, que também foi utilizado nessa pesquisa. Ademais, para todos os cenários testados na literatura indicada, a ML1 demonstrou ser resistente a *outliers*, apresentando, assim, resíduos absolutos relativamente maiores para os *outliers* das redes analisadas, mesmo em cenários de múltiplos *outliers*.

Entretanto, não há na literatura uma investigação que valide a escolha do resíduo absoluto do ajustamento como parâmetro de comparação para identificação de *outliers*. Uma vez que o ajustamento pela ML1 adota determinado modelo estocástico, pode-se imaginar que o resíduo absoluto ponderado (pelo peso da respectiva observação nesse modelo estocástico) seja também um parâmetro viável de comparação. É ainda passível de investigação a potencial eficácia de um ajustamento com pesos unitários para as observações na identificação de *outliers*, para o caso do resíduo absoluto ser mesmo o parâmetro a ser analisado. Essa questão foi abordada nos experimentos dessa pesquisa.

Ademais, não existe um critério objetivo estabelecido para classificação de uma observação como *outlier* a partir de seu respectivo resíduo absoluto no ajuste pela ML1. É razoável visualizar que isto se torna crítico, principalmente para *outliers* de magnitude relativamente menores e para caracterização de uma rede como sem *outlier*. O Classificador VL1 proposto nessa pesquisa, que será visto no Capítulo 5, estabelece possíveis critérios a serem adotados.

Contudo, pode surgir o seguinte questionamento: dado que a ML1 provê uma estimação dos parâmetros mais robusta que a do MMQ, por que não utilizá-lo como o próprio método de estimação da rede, em substituição ao MMQ? VALERO e MORENO (2005) mostram que, de fato, para o caso de ocorrência de *outliers*, o ajustamento pela ML1 seria o mais apropriado. Porém, na ausência deles e ajuste pelo MMQ, além da solução única, obtém-se a estimativa ótima de mínima variância para os parâmetros e de máxima verossimilhança, considerando a normalidade dos erros das observações. Assim, é mais interessante que a ML1 auxilie na identificação de *outliers*, visando um posterior ajuste da rede pelo MMQ (MARSHALL e BETHEL, 1996), para o qual, também, a teoria de propagação de variâncias (essencial na análise do ajustamento de dados geodésicos) é bem definida na literatura.

Além disso, a ML1 apresenta características que dificultam sua aceitação no ajustamento de redes geodésicas. Como visto na seção 2.1, sua solução ótima, em geral, não é única, o que

pode trazer problemas para a repetibilidade e caracterização dos resultados do ajustamento. Ademais, seus resultados apresentam resíduo nulo para parte de observações, o que, na prática, significa considerar a inexistência de erros aleatórios nas mesmas e a superestimação de seus pesos (pois passam a configurar injunções na rede).

3.3 INFLUÊNCIA DA GEOMETRIA DA REDE

O conteúdo apresentado nessa seção está baseado no trabalho de HEKIMOGLU *et al.* (2011). Em redes geodésicas, falhas na identificação de *outliers* podem ocorrer devido a pouca redundância das respectivas observações (o que os autores chamam de fraca geometria da rede ou rede com geometria não confiável), independentemente do método utilizado na identificação (classificador) deles. Nesse contexto, é fundamental que na pré-análise referente ao projeto da rede, seja essa concebida com geometria confiável, que venha a permitir uma posterior identificação de *outliers*, caso ocorram, no processamento da mesma.

Assim, foi proposto um método para identificação de redundância deficiente na geometria de redes geodésicas, pela contagem de equações medianas das observações: o Método das Equações Medianas (MEM). Equações medianas (EM) de uma observação são equações de condição independentes, organizadas para verificação da resistência a *outliers* da própria observação, a qual também é considerada uma equação mediana dela mesma. Cada observação da rede pode aparecer somente 1 vez, no máximo, no conjunto de equações medianas de uma observação.

Um conceito relevante a ser entendido é o de ponto de quebra de um estimador. Diz respeito ao percentual limite de *outliers* nas observações a serem ajustadas, até o qual a ocorrência deles não contamina a estimação em si. Para o MMQ, um *outlier* é suficiente para degradar a estimação, assim, seu ponto de quebra é zero. Já a mediana, que corresponde à norma L1 mínima para o caso unidimensional, possui ponto de quebra mais alto. Apenas se um percentual maior ou igual a 50% da redundância de uma observação for composta por *outliers*, sua estimação não mais será resistente a eles. O MEM faz uso desta propriedade para verificar a confiabilidade da geometria em trechos da rede. Ressalta-se que essa aplicação da norma L1 é diferente do ajustamento pela ML1 visto no Capítulo 2. Aqui se está explorando uma propriedade da norma L1 no caso unidimensional com a única finalidade de analisar a resistência da rede a *outliers*. Não se trata nem de ajustamento das observações, nem de um método de identificação de *outliers*.

Pelo MEM, se o número de equações medianas de uma observação \mathbf{h}_i for maior ou igual a 3, esta é controlável contra 1 *outlier* na rede. Assim, nesse cenário, a identificação da observação como *outlier* por algum classificador, caso ocorra, pode ser confirmada. Caso contrário (*outlier* identificado em observação não controlável), o resultado da identificação não pode ser confirmado de imediato, devido à fraca geometria da rede.

A FIG. 3.5a mostra uma rede de nivelamento em que todas as observações são confiáveis contra 1 *outlier*, pois possuem 3 equações medianas (TAB. 3.1), sendo $\mathbf{h}_i^{(j)}$ a j -ésima equação mediana da i -ésima observação. Assim, uma possível identificação de *outlier* em uma das observações da rede seria de possível confirmação. Porém, na rede da FIG 3.5b, somente \mathbf{h}_5 é controlável. As observações \mathbf{h}_1 a \mathbf{h}_4 não o são, pois possuem apenas 2 equações medianas cada (TAB 3.2). Neste caso, somente a identificação de *outlier* em \mathbf{h}_5 seria confirmável.

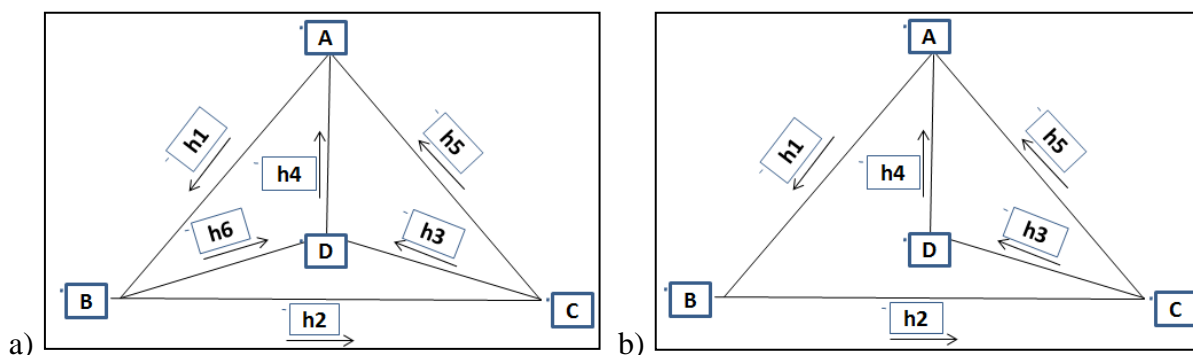


FIG. 3.5: Redes de nivelamento com e sem geometria confiável contra 1 *outlier*.

TAB. 3.1: Equações medianas das observações da rede da FIG. 3.5a.

\mathbf{h}_1	\mathbf{h}_2	\mathbf{h}_3	\mathbf{h}_4	\mathbf{h}_5	\mathbf{h}_6
$\mathbf{h}_1^{(1)}=\mathbf{h}_1$	$\mathbf{h}_2^{(1)}=\mathbf{h}_2$	$\mathbf{h}_3^{(1)}=\mathbf{h}_3$	$\mathbf{h}_4^{(1)}=\mathbf{h}_4$	$\mathbf{h}_5^{(1)}=\mathbf{h}_5$	$\mathbf{h}_6^{(1)}=\mathbf{h}_6$
$\mathbf{h}_1^{(2)}=-\mathbf{h}_5-\mathbf{h}_2$	$\mathbf{h}_2^{(2)}=-\mathbf{h}_1-\mathbf{h}_5$	$\mathbf{h}_3^{(2)}=\mathbf{h}_5-\mathbf{h}_4$	$\mathbf{h}_4^{(2)}=-\mathbf{h}_3+\mathbf{h}_5$	$\mathbf{h}_5^{(2)}=\mathbf{h}_3+\mathbf{h}_4$	$\mathbf{h}_6^{(2)}=-\mathbf{h}_1-\mathbf{h}_4$
$\mathbf{h}_1^{(3)}=-\mathbf{h}_4-\mathbf{h}_6$	$\mathbf{h}_2^{(3)}=\mathbf{h}_6-\mathbf{h}_3$	$\mathbf{h}_3^{(3)}=-\mathbf{h}_2+\mathbf{h}_6$	$\mathbf{h}_4^{(3)}=-\mathbf{h}_6-\mathbf{h}_1$	$\mathbf{h}_5^{(3)}=-\mathbf{h}_2-\mathbf{h}_1$	$\mathbf{h}_6^{(3)}=\mathbf{h}_2+\mathbf{h}_3$

TAB. 3.2: Equações medianas das observações da rede do FIG 3.5b.

\mathbf{h}_1	\mathbf{h}_2	\mathbf{h}_3	\mathbf{h}_4	\mathbf{h}_5
$\mathbf{h}_1^{(1)}=\mathbf{h}_1$	$\mathbf{h}_2^{(1)}=\mathbf{h}_2$	$\mathbf{h}_3^{(1)}=\mathbf{h}_3$	$\mathbf{h}_4^{(1)}=\mathbf{h}_4$	$\mathbf{h}_5^{(1)}=\mathbf{h}_5$
$\mathbf{h}_1^{(2)}=-\mathbf{h}_5-\mathbf{h}_2$	$\mathbf{h}_2^{(2)}=-\mathbf{h}_1-\mathbf{h}_5$	$\mathbf{h}_3^{(2)}=\mathbf{h}_5-\mathbf{h}_4$	$\mathbf{h}_4^{(2)}=-\mathbf{h}_3+\mathbf{h}_5$	$\mathbf{h}_5^{(2)}=\mathbf{h}_3+\mathbf{h}_4$
---	---	---	---	$\mathbf{h}_5^{(3)}=-\mathbf{h}_2-\mathbf{h}_1$

Ainda sobre a geometria da rede da FIG 3.5b, nota-se ser impossível construir, por exemplo, uma nova equação mediana para \mathbf{h}_1 (visando completar o total de 3 equações medianas) sem a participação de \mathbf{h}_2 . Isto denota grande dependência entre as mesmas e, assim, um possível *outlier* em \mathbf{h}_1 poderia contaminar \mathbf{h}_2 , e vice-versa. Ou seja: caso um classificador indicasse a presença de *outlier* em \mathbf{h}_1 ou \mathbf{h}_2 , apesar de se saber a ocorrência do *outlier*, seria impossível afirmar qual das 2 observações (\mathbf{h}_1 ou \mathbf{h}_2) é, de fato, um *outlier*.

Para múltiplos *outliers* na rede, equações medianas especiais, que são equações de condição diferentes entre si, que não incluem os *outliers*, mas admitem a repetição de observações em seu conjunto, também são admitidas na contagem de equações medianas. Os autores, nesse caso, chamam de equações medianas estendidas o conjunto de equações medianas e de equações medianas especiais de uma observação. Contudo, a estratégia do MEM para múltiplos *outliers* é similar ao caso de um único: a observação é controlável se sua quantidade de equações medianas com *outliers* for inferior a 50% do total de equações medianas estendidas da respectiva observação, ou seja, não ultrapasse o ponto de quebra da mediana.

Sejam \mathbf{h}_1 e \mathbf{h}_4 *outliers* simultâneos na rede da FIG 3.6. A TAB 3.3 apresenta, por exemplo, as equações medianas estendidas dessas 2 observações e de \mathbf{h}_8 . Todas possuem *outliers* em 2 de 3 de suas equações medianas. Para \mathbf{h}_1 e \mathbf{h}_4 foi possível formar 2 equações medianas especiais para elas. Assim, com um total de 5 equações medianas estendidas e somente duas delas com *outliers*, a mediana pode distinguir os *outliers* para \mathbf{h}_1 e \mathbf{h}_4 . Entretanto, apesar de \mathbf{h}_8 não ser *outlier*, não é possível formar equações medianas especiais para \mathbf{h}_8 que não incluam \mathbf{h}_1 e \mathbf{h}_4 (*outliers*). Com isso, *outliers* simultâneos em \mathbf{h}_1 e \mathbf{h}_4 não são controláveis, pois contaminam \mathbf{h}_8 . Ou seja, caso um classificador indicasse \mathbf{h}_1 e \mathbf{h}_4 como *outliers*, o resultado não poderia ser confirmado de imediato.

Para essa rede da FIG 3.6, de um total de 171 pares possíveis de *outliers*, 117 são controláveis e 54 não são controláveis. Maiores informações e exemplos detalhados no trabalho de referência utilizado nessa seção.

Assim sendo, sintetizando, tem-se que:

- a) no caso de somente um *outlier*, a contagem das equações medianas daquela específica observação é suficiente para conclusão sobre a resistência (ou não) da rede a ele;
- b) no caso de múltiplos *outliers*, a análise deve ser realizada para todas as observações da rede, pois cada uma pode não ser resistente aos *outliers*, mesmo que elas próprias não sejam *outliers*.

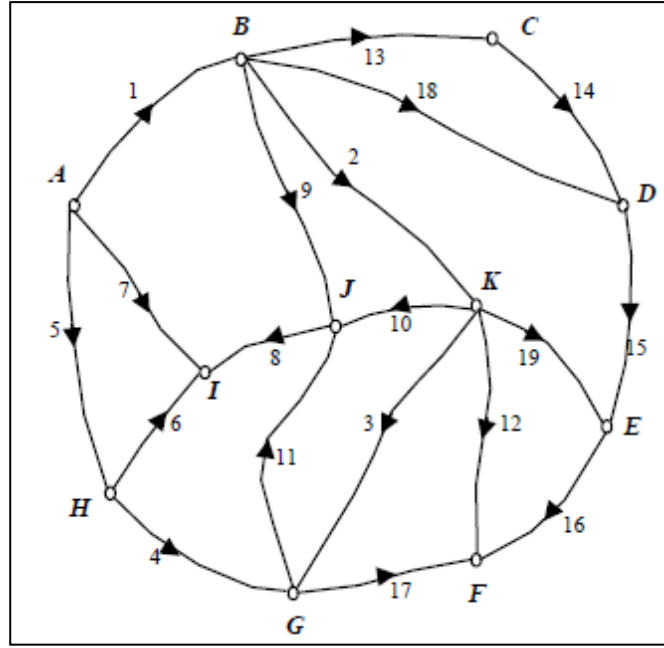


FIG. 3.6: Rede analisada contra 2 *outliers*.
Fonte: HEKIMOGLU *et al.* (2011)

TAB. 3.3: Equações medianas estendidas.

h_1	h_4	h_8
EM	EM	EM
$h_1^{(1)}=h_1$	$h_4^{(1)}=h_4$	$h_8^{(1)}=h_8$
$h_1^{(2)}=h_7-h_8-h_9$	$h_4^{(2)}=h_6-h_8-h_{11}$	$h_8^{(2)}=h_7-h_1-h_9$
$h_1^{(3)}=-h_2-h_3+h_4+h_5$	$h_4^{(3)}=-h_5+h_1+h_2+h_3$	$h_8^{(3)}=h_6-h_4-h_{11}$
EM Especiais	EM Especiais	EM Especiais
$h_1^{(4)}=-h_5+h_6-h_8-h_9$	$h_4^{(4)}=-h_5+h_7-h_8-h_{11}$	-----
$h_1^{(5)}=h_7-h_8-h_{10}-h_2$	$h_4^{(5)}=h_6-h_8-h_{10}+h_3$	-----

3.4 AVALIAÇÃO DOS CLASSIFICADORES DE *OUTLIERS*

Na identificação de *outliers* em redes geodésicas, é usual que a comparação do desempenho de classificadores seja realizada pelas respectivas taxas de sucesso (TS) (EQ 3.6) nas redes analisadas. Tal abordagem para a avaliação de classificadores de *outliers*, que pode ser vista, por exemplo, em (KLEIN *et al.*, 2015) e (HEKIMOGLU e ERENOGLU, 2007), foi adotada também nessa pesquisa.

$$TS = \frac{\text{quantidade de classificações corretas}}{\text{total de experimentos}}, 0 \leq TS \leq 1 \quad (\text{EQ 3.6})$$

Nesse contexto, torna-se necessária a obtenção de inúmeros cenários de rede para análise. O estado da arte para simulação desses cenários em redes geodésicas é a aplicação do Método de Monte Carlo, também chamado de Técnica de Monte Carlo ou Simulação de Monte Carlo, assunto da seção seguinte.

3.4.1 SIMULAÇÃO DE MONTE CARLO APLICADA

A Simulação de Monte Carlo (SMC) é um tipo de simulação que se baseia em repetidas amostras aleatórias das variáveis de entrada, permitindo dessa forma uma abrangente caracterização da variação dos resultados do modelo a ser estudado. É pressuposto conhecer a distribuição de probabilidade associada ao conjunto de variáveis (RAYCHAUDHURI, 2008). Quanto maior a quantidade de amostras aleatórias simuladas, mais acurada será a referida caracterização. Desenvolvida no contexto da construção da bomba atômica em 1944, atualmente é utilizada em diversas áreas, como finanças (análise de risco) e gestão de projetos.

Entretanto, sua aplicação tem se tornado usual também na avaliação de classificadores (métodos de identificação) de *outliers* em redes geodésicas, podendo ser vista, por exemplo, em (KLEIN *et al.*, 2015), (HEKIMOGLU e ERENOGLU, 2007) e (DURDAG *et al.*, 2016). Apesar dessas referências terem aplicado SMC considerando a *regra dos 3σ para o erro grosseiro*, nos experimentos dessa pesquisa essas foram adaptadas para a *regra dos 3σ para o erro total*, conforme discussão da seção 3.1.

A seguir faz-se uma caracterização da literatura indicada (utilizando a diferenciação dos conceitos de *outlier* e erro grosseiro estabelecida na seção 3.1), evidenciando elementos relevantes para essa pesquisa, a saber:

- (KLEIN *et al.*, 2015)
 - a) emprego de 2 redes GNSS sem *outliers*, obtidas a partir de dados reais;
 - b) erros grosseiros de magnitudes (com distribuição uniforme) entre $3\sigma_i$ e $9\sigma_i$ (positivos ou negativos), sendo σ_i o DP da respectiva *i-ésima* observação, foram adicionados a observações da rede, as quais supostamente passaram a configurar *outliers*;
 - c) da análise da TS de diferentes classificadores em 15.000 cenários simulados (5.000 com 1 erro grosseiro, 5.000 com 2 erros grosseiros e 5.000 com 3 erros grosseiros), para cada uma das 2 redes analisadas, foi possível caracterizar e comparar o desempenho deles;
 - d) inserção dos erros grosseiros diretamente em observações que continham erros aleatórios.

- (HEKIMOGLU e ERENOGLU, 2007) e (DURDAG *et al.*, 2016)

a) além dos erros grosseiros, as próprias observações da rede foram também simuladas via SMC. Para tal, partindo-se de redes propositalmente “perfeitas” simuladas (sem erros de medida, nem mesmo aleatórios), a estratégia foi simular repetidas amostras da respectiva rede a partir dos σ_i conhecidos das observações com distribuição normal. Após isso, veio a inserção proposital dos erros grosseiros (para que as respectivas observações viessem a configurar *outliers*) de forma parecida com a feita por KLEIN *et al.* (2015), com algumas variações nos intervalos de magnitudes deles;

b) a principal diferença em relação à KLEIN *et al.* (2015) foi a retirada do erro aleatório da observação antes da inserção do erro grosseiro. Obviamente, isto somente foi possível porque se tratavam de observações simuladas, nas quais se podia determinar exatamente o erro aleatório, algo inviável com observações reais;

c) sendo **m** a quantidade de observações da rede, foram simuladas **m*100** “boas” amostras da rede (redes com erros aleatórios, mas sem *outliers*). Em seguida, para cada uma dessas, **100** cenários foram criados com um erro grosseiro (com distribuição uniforme, dentro de intervalos especificados múltiplos de σ_i) inserido aleatoriamente nas observações, totalizando **m*10.000** (**m*100*100**) cenários para cada rede “perfeita” analisada;

d) análise de cenários nos quais todas as observações têm a mesma variância (homocedasticidade), cenários em que as observações têm variâncias diferentes (heterocedasticidade) e cenários em que há uma mistura dos dois anteriores (heterogeneidade). Esse ponto foi específico da investigação daqueles autores, porém nessa pesquisa as observações serão simuladas sempre com ocorrência de heterocedasticidade, caso geral das redes geodésicas.

4 EXTENSÃO DO CONCEITO DE ESTIMAÇÃO ROBUSTA E CONTEXTO DA APLICAÇÃO DA NORMA L_∞ EM REDES ALTIMÉTRICAS

Há autores, como YU *et al.* (2017), que, ao utilizarem um método robusto no ajustamento das observações, consideram que foi realizada uma estimação robusta da rede (lado esquerdo da FIG. 4.1). Tal interpretação pode ser reforçada pelo fato de que a aplicação de métodos robustos, além de diminuir ou isolar o efeito de possíveis *outliers*, resulta também na própria estimação dos parâmetros incógnitos, como no caso da ML1 na FIG 1.1, o que a torna aceitável.

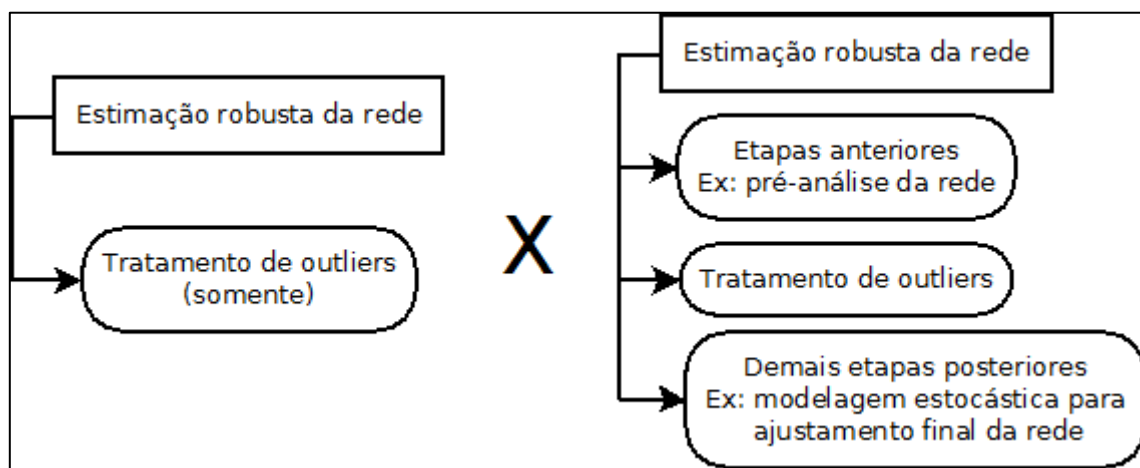


FIG. 4.1: Interpretações para a estimação robusta de redes.

Por outro lado, é razoável considerar que a estimação robusta de uma rede geodésica não se limita à utilização de métodos robustos no seu ajustamento. Ora, uma pré-análise (planejamento) adequada da rede, a atenção às especificações do trabalho no levantamento de campo e o tratamento de efeitos sistemáticos nas observações, por exemplo, não fariam parte, também, de um processo de estimação robusta? Estas atividades não estariam contribuindo para a resistência da rede a *outliers* e para a qualidade dos resultados da estimação?

Apesar do conceito matemático de robustez estar diretamente relacionado à resistência a desvios em relação às premissas, a robustez em si, num contexto mais amplo, faz referência à qualidade de vigoroso e forte (MICHAELIS, 2018). Neste sentido, o tratamento de *outliers* pode ser visto como uma das possíveis etapas, dentre outras, de um processo mais completo: a estimação robusta da rede, cujo objetivo (mais geral) seria garantir resultados “fortes”, de qualidade “vigorosa” para a rede, o que incluiria também, mas não exclusivamente, a questão da resistência a *outliers*. Nesse caso, a estimação robusta englobaria qualquer técnica para o

tratamento de *outliers* na rede, além de outras para as demais etapas do processo (lado direito da FIG. 4.1). Até mesmo a investigação de um modelo estocástico mais apropriado para aplicação no ajustamento final (posterior ao tratamento de *outliers*) da rede pelo MMQ estaria incluída no contexto da estimação robusta. Essa última interpretação foi a adotada nesta pesquisa.

Assim, esse capítulo abordou aspectos da estimação robusta de redes altimétricas, com exceção do tratamento de *outliers*, já visto no capítulo anterior. Foi dada ênfase na modelagem estocástica para ajustamento final da rede pelo MMQ, contexto da aplicação da norma L_∞ nessa pesquisa.

4.1 CONCEITOS RELACIONADOS AO LEVANTAMENTO ALTIMÉTRICO

Levantamento altimétrico é aquele que objetiva a determinação de alturas e/ou altitudes relativas a uma superfície de referência, valendo-se de medidas da diferença de nível entre pontos do terreno (não serão feitas maiores considerações a nível teórico ou prático sobre os termos, por fugir ao escopo da pesquisa). Essas medidas podem ser obtidas por métodos diretos ou indiretos. Os métodos de levantamento mais precisos utilizados na implantação de redes altimétricas são o nivelamento geométrico (direto) e o trigonométrico (indireto).

Entre métodos existentes, todos indiretos, estão o barométrico, baseado em diferenças de pressão atmosférica, os aéreos ou orbitais, baseados no tempo que a radiação eletromagnética emitida (normalmente radar ou *laser*) leva para ir e voltar ao sensor embarcado, após interação com a superfície, e o nivelamento a partir de dados de Sistemas Globais de Navegação por Satélite. Estações *laser* terrestres também são aplicadas em levantamentos altimétricos locais, normalmente em apoio à engenharia civil.

No nivelamento geométrico, a norma da Associação Brasileira de Normas Técnicas (ABNT) que trata do tema coloca que os desníveis são computados “por intermédio de leituras correspondentes a visadas horizontais, obtidas com um nível, em miras colocadas verticalmente nos referidos pontos” (ABNT, 1994). Os níveis são equipamentos que definem o plano ortogonal ao vetor da gravidade no local, possibilitando realizar a leitura de nível (visada) nas miras (régua adaptada). Um lance é a medida do desnível entre duas miras (ré e vante) verticais (FIG. 4.2). O nivelamento geométrico é o padrão imposto em IBGE (2017) para a implantação de redes altimétricas a nível nacional.

Já no nivelamento trigonométrico, diferenças de nível são calculadas por relações trigonométricas, a partir das medidas (obtidas com teodolito ou estação total) da distância e

do ângulo vertical da direção entre os pontos (GUILANI, 2010). Apesar de menos preciso, é mais flexível que o geométrico, por permitir visadas inclinadas, algo especialmente útil em áreas de difícil instalação dos níveis, como encostas íngremes.

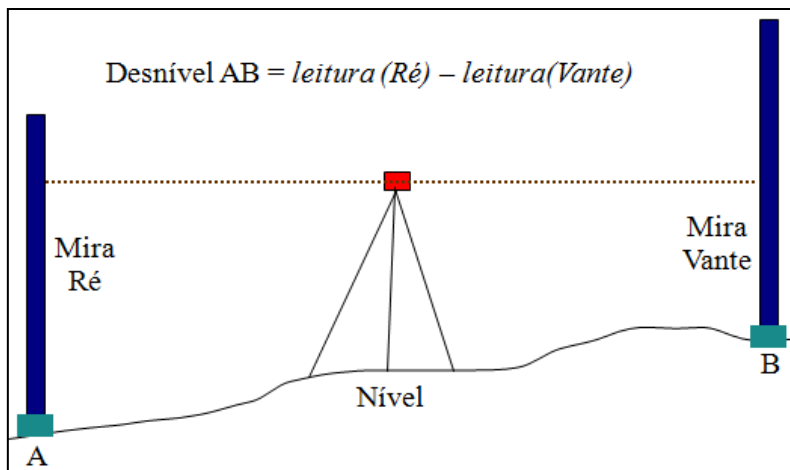


FIG. 4.2: Lance de nivelamento geométrico.

Em redes altimétricas, diversos lances em sequência, normalmente entre duas referências de nível (RRNN), formam uma seção, cujo desnível é a soma algébrica dos desníveis de seus lances. Linhas de nivelamento se referem normalmente a conjuntos de seções em sequência entre 2 RRNN principais (normalmente aquelas com no mínimo 3 conexões). Um circuito de nivelamento corresponde a um conjunto de linhas que forma uma poligonal fechada e uma rede de nivelamento é uma rede na qual as observações são linhas de nivelamento, formando uma malha de circuitos justapostos. Exemplos desses conceitos na rede de nivelamento da FIG 4.3.

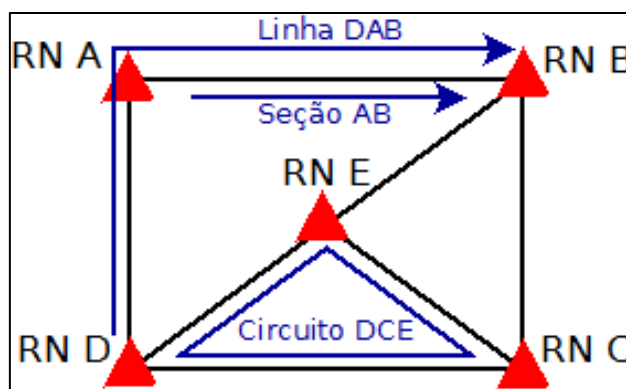


FIG. 4.3: Exemplos de componentes de uma rede altimétrica.

4.2 ASPECTOS RELEVANTES PARA A ESTIMAÇÃO ROBUSTA

O conteúdo dessa seção aborda aspectos relevantes para a estimação robusta de redes altimétricas. Não fazem parte das hipóteses a serem investigadas nessa pesquisa, porém são relevantes para entendimento do posicionamento delas na estimação robusta de redes altimétricas. Foram apontados conceitos importantes sobre planejamento (pré-análise), procedimentos de campo, tratamento de efeitos sistemáticos e de reverberação em redes altimétricas, legislação nacional relacionada, confiabilidade interna e externa e análise de robustez. Entretanto, por ser o conteúdo introdutório e apenas para situar o posicionamento das hipóteses da pesquisa no contexto de uma estimação robusta de rede altimétrica, há necessidade de um maior aprofundamento para o leitor interessado.

A primeira etapa para implantação de uma rede geodésica é o seu planejamento. Além de detalhes logísticos e administrativos, nele deve-se realizar uma análise técnica, um estudo prévio da rede para “definir as especificações do levantamento de campo a fim de atender aos pré-requisitos desejados em termos de qualidade” (GEMAEL *et al.*, 2015). Sugere-se a consulta do conteúdo relacionado na referida obra para a pré-análise de uma rede altimétrica. Ademais, conforme visto na seção 3.3, a análise de geometria pelo MEM também deve ser considerada no planejamento da rede num processo de estimação robusta. Entretanto, devido ao custo da coleta de observações, as redes altimétricas tendem a ter geometria menos resistente a *outliers*.

As especificações e normas gerais para levantamentos altimétricos a nível nacional estão contidas em IBGE (2017), que adota o nivelamento geométrico como padrão para implantação de redes altimétricas a nível nacional, conforme visto na seção 4.1. Vale ressaltar alguns procedimentos de campo especificados nesta norma, fundamentais para uma estimação robusta da rede. Dentre os procedimentos, ressalta-se: a) adoção de distância das visadas de ré e vante aproximadamente iguais nos respectivos lances (considerando 60 m a ideal) para minimização de efeitos da curvatura terrestre e da refração atmosférica; b) realização de leituras 50 cm acima do solo para se evitar turbulências da reverberação. Entre outros detalhes, tal legislação trata ainda de comprimentos de circuitos e linhas, condições dos instrumentos de medição e controle de qualidade.

Dentre outros detalhes relevantes, citados também em ABNT (1994), estão a alternância de miras nas posições de ré e vante e a utilização chapas, pinos e sapatas para que as miras não sejam apoiadas sobre o solo.

Quanto aos efeitos sistemáticos nas observações, além das já citadas precauções técnicas em campo a serem tomadas, recomenda-se a abordagem de GHILANI (2010) para o tratamento dos efeitos sistemáticos de colimação, curvatura da Terra e refração atmosférica no nivelamento geométrico. A mesma obra também trata dos efeitos no nivelamento trigonométrico.

Os aspectos citados até agora são executados antes do tratamento de *outliers*. Posteriores a ele, sendo após até mesmo do ajustamento final da rede pelo MMQ, estão ainda a teoria da confiabilidade, que trata das confiabilidades interna e externa da rede (BAARDA, 1968), e a análise de robustez (BERBER, 2006). Considerando que a identificação de *outliers* foi feita pelo DS, servem para determinar os mínimos erros detectáveis, o efeito desses nos parâmetros estimados e a resistência da rede a eles, respectivamente. Mais detalhes nas obras mencionadas. Esses aspectos podem também ser considerados ainda na fase de planejamento da rede, para que sua confiabilidade externa não extrapole determinado valor, por exemplo.

4.3 MODELO ESTOCÁSTICO USUAL EM REDES ALTIMÉTRICAS

O modelo estocástico mais usual adota pesos proporcionais ao inverso do comprimento das respectivas linhas em redes de nivelamento geométrico, ou seja, nele se considera que os erros aleatórios do nivelamento se propagam com a raiz quadrada do comprimento da linha. Entretanto, outras abordagens para o modelo estocástico são possíveis, apesar de ainda pouco exploradas na literatura. Um exemplo de menção a outra possível abordagem foi feita por TORGE (2011), que coloca a propagação de erros aleatórios como proporcionais à raiz quadrada do número de estações da linha.

Para o modelo mais usual citado, GHILANI (2010) apresenta uma demonstração matemática, a qual aceita algumas conjecturas, que justifica tal abordagem. Essa demonstração será revisitada a seguir, enfatizando tais conjecturas, pois devem ser sempre verificadas se, de fato, correspondem à realidade da rede a ser ajustada, antes de se atribuir o peso das observações por esta abordagem usual. Para informações introdutórias sobre modelo estocástico em redes geodésicas recomenda-se (GEMAEL *et al.*, 2015). Vale ressaltar que, apesar de justificada apenas para redes de nivelamento geométrico, esta abordagem para o modelo estocástico é por vezes aplicada, na prática geodésica, para redes altimétricas em geral.

O método das visadas iguais (MVI) é o padrão imposto em IBGE (2017) e ABNT (1994) para o nivelamento geométrico na implantação de redes de nivelamento. Nele, as distâncias

das visadas (ré e vante) num lance são iguais ($\mathbf{D}_{ré}=\mathbf{D}_{vante}=\mathbf{D}$), assim como seus respectivos DP, o que minimiza os efeitos da curvatura terrestre, da refração atmosférica e dos erros de colimação do nível.

Primeiramente, GHILANI (2010) demonstra, inclusive com exemplo numérico, que o erro estimado numa linha \mathbf{e}_{cm} causado por efeitos de erros de calagem da mira (EQ 4.1, na qual $\Delta\mathbf{Elev}$ é a variação de nível na linha e β o ângulo de descentragem da bolha da mira) é desprezível (Conjectura I), considerando as mesmas bem ajustadas e que as bolhas sejam cuidadosamente centradas em cada visada, ignorando-o nos cálculos de nivelamento geométrico seguintes.

$$e_{cm} = \frac{\Delta Elev * \sin(\beta)}{2} \quad (\text{EQ 4.1})$$

A partir do MVI, considera que também as distâncias das visadas dos diversos lances (não somente as de um mesmo lance) de uma linha são iguais (Conjectura II). Assim, aplicando a propagação de variâncias para modelagem dos efeitos dos erros aleatórios no nivelamento, a formulação da EQ 4.2 é apresentada para o DP do desnível em uma linha, sendo \mathbf{N} o número de lances, $\sigma_{r/D}$ o erro estimado na leitura da mira por unidade de comprimento da visada, σ_{α} o da colimação do nível, σ_D o erro estimado na medida das distâncias das visadas, \mathbf{C} o efeito da curvatura terrestre, \mathbf{R} da refração atmosférica e α da colimação.

$$\sigma_{\Delta h} = \sqrt{2ND^2(\sigma_{r/D}^2 + \sigma_{\alpha}^2) + 2N\sigma_D^2(\alpha + \frac{CR(D)}{500000})^2} \quad (\text{EQ 4.2})$$

A seguir é demonstrado, com exemplo numérico, que a magnitude do segundo termo da radiciação da EQ 4.2 (parte de $\sigma_{\Delta h}$ causada pelo erro na medida de distância das visadas) é desprezível (Conjectura III), com exceção a trabalhos de maior precisão. Com isso, a equação para a variância de uma linha no nivelamento geométrico pelo método das visadas iguais passa a ser dada pela EQ 4.3.

$$\sigma_{\Delta h}^2 = 2ND^2(\sigma_{r/D}^2 + \sigma_{\alpha}^2) \quad (\text{EQ 4.3})$$

Dada a Conjectura II, e sendo \mathbf{l}_i o comprimento total da linha, o número de lances da respectiva linha pode ser expresso pela EQ 4.4. Ademais, substituindo esta na EQ 4.3, se obtém a EQ 4.5.

$$N = \frac{l_i}{2D} \quad (\text{EQ 4.4})$$

$$\sigma_{\Delta h}^2 = l_i D (\sigma_{r/D}^2 + \sigma_{\alpha}^2) \quad (\text{EQ 4.5})$$

Prosseguindo, além de \mathbf{D} já ter sido considerada constante em uma linha (pela Conjectura II), se aceita que também o é em toda a rede (Conjectura IV). Ademais $\sigma_{r/D}$ (Conjectura V) e

σ_a (Conjectura VI) também são dados como constantes, o que possibilita que a EQ 4.5 seja reescrita pela EQ 4.6, sendo k uma constante. Assim, os pesos p_i (inverso da variância) das linhas de uma rede passam a ter a forma da EQ 4.7, e, como esses são relativos, esta última pode ser simplificada pela EQ 4.8, que os expressa como inversamente proporcionais ao comprimento das respectivas linhas.

$$\sigma_{\Delta h}^2 = l_i k \quad (\text{EQ 4.6})$$

$$p_i = \frac{1}{l_i k} \quad (\text{EQ 4.7})$$

$$p_i = \frac{1}{l_i} \quad (\text{EQ 4.8})$$

Por fim, deve-se atentar que os pesos têm a unidade do inverso do desvio padrão (que tem unidade do sistema métrico) ao quadrado. Ou seja, os pesos têm a unidade adotada para os desvios padrão da rede elevada a “-2”. Apesar de proporcionais ao inverso do comprimento de suas respectivas linhas, eles não passam a ter a unidade dele.

Sobre as conjecturas aceitas na demonstração de GUILANI (2010), resumidas na TAB 4.1, os seguintes comentários são colocados:

TAB. 4.1: Conjecturas do modelo estocástico usual

Conjectura:	Descrição
I	Efeitos de erros de calagem da mira são desprezíveis
II	Distância das visadas são constantes em cada linha de nivelamento
III	Segundo termo da radiciação da EQ 4.2 é desprezível
IV	Distância das visadas são constantes em toda a rede de nivelamento
V	Erro na leitura da mira por unidade de comprimento da visada é constante em toda a rede de nivelamento
VI	Erro de colimação do nível é constante em toda a rede de nivelamento

a) quanto à I, da própria EQ 4.1 verifica-se que o efeito do erro de calagem da mira cresce diretamente com o desnível da linha. No exemplo numérico apresentado por GUILANI (2010), para um desnível de 22,865 m, o erro estimado foi de 0,02 mm. Porém, em linhas de maior desnível, como 500 m, o erro subiria (com o equipamento do exemplo) para 0,44 mm, valor que provavelmente não seria desprezível, dependendo da precisão do nivelamento. Ademais, a própria premissa dessa conjectura de que as miras estão bem ajustadas e que são cuidadosamente caladas em todos os lances da rede pode não ocorrer na prática. O “bom” ajuste das miras pode ser degradado, por exemplo, pelo desgaste das mesmas em condições

extremas (normais a nível nacional) de temperatura por longos períodos no campo. Até mesmo a precisão do operador na calagem de bolhas pode ser prejudicada, por serviços exaustivos em regiões de difícil acesso, como áreas de floresta;

b) quanto à II, IBGE (2017) propõe que o comprimento máximo das visadas seja de 60 m. Entretanto, é de se esperar que, para parte das visadas de uma rede, isso seja inviabilizado no terreno, por exemplo, por relevo acidentado ou obstáculos naturais, como rios cruzando a linha de nivelamento. A própria legislação citada aceita que isto não seja estritamente seguido, desde que o comprimento da visada seja mantido num mesmo lance;

c) quanto à III, o exemplo numérico mostrado pelo autor indica fortemente que tal conjectura é plenamente aceitável;

d) quanto à IV, trata-se de uma extensão da Conjectura II (a qual já é de difícil garantia de ocorrência na prática) para toda a rede; e

e) quanto à V e VI, se considerou que todas as linhas da rede sejam levantadas com equipamentos de igual precisão. Porém, nos levantamentos de maiores dimensões, comuns a nível nacional, é usual que várias linhas sejam medidas simultaneamente (obviamente, com diferentes equipamentos). Assim, pode ocorrer de diferentes modelos de níveis e miras (com precisões distintas para a medição dos desníveis) serem utilizados em diferentes linhas. Some-se a isso a questão de que, mesmo que se utilizem equipamentos do mesmo modelo, suas precisões podem não ser mais iguais após anos de uso e, até mesmo, possíveis calibrações recentes dos mesmos podem ser degradadas no decorrer do serviço, por condições de desgaste extremas.

Concluindo, a usual abordagem do peso pelo inverso do comprimento da linha em redes de nivelamento geométrico é demonstrável. Entretanto, antes de uma possível decisão de adotá-la na estimação de sua rede altimétrica, é interessante que o geodesta atente para as conjecturas que são feitas em tal abordagem. Conforme visto, a garantia de ocorrência na prática delas é pouco viável, o que justifica a busca de abordagens alternativas para o modelo estocástico de redes altimétricas, como feito nessa pesquisa.

4.4 UMA ABORDAGEM ALTERNATIVA PARA O MODELO ESTOCÁSTICO

Voltando ao exemplo da FIG. 1.1, nele foi visto que o ajustamento pela ML_{∞} causou a máxima distribuição dos resíduos do *outlier*. Devido a essa característica desse ajustamento, para o caso de *outliers* na amostra de observações ele não é apropriado, pois a distribuição de resíduos causada tende a mascarar os *outliers* (ABDELMALEK e MALEK, 2008).

Entretanto, a ML_{∞} acarretou também, de maior interesse agora, justamente devido à distribuição de resíduos, o mínimo MRA (2,5 u.m.) dentre os métodos de ajustamento (TAB. 4.2), o que tornou também os resíduos mais homogêneos em relação ao MMQ e à ML_1 . Com isso, após o tratamento dos *outliers* da rede, visando aproveitar tal característica da ML_{∞} , é passível de investigação se seria vantajosa, para a qualidade do ajustamento pelo MMQ, a construção de um modelo estocástico alternativo com pesos das observações que minimizem seu MRA.

TAB. 4.2: Máximo resíduo absoluto no ajustamento da FIG 1.1

Métodos de ajustamento	ML_1	MMQ	ML_{∞}
MRA	5,0 u.m.	4,0 u.m.	2,5 u.m.

Assim, a abordagem alternativa para o modelo estocástico do ajustamento pelo MMQ de redes altimétricas dessa pesquisa visou minimizar o MRA desse ajustamento. Para cálculo do mínimo MRA, foi aplicado um pré-ajuste da rede pela ML_{∞} com pesos iguais para as observações. O detalhamento do procedimento para construção de tal modelo estocástico será visto na seção 5.4.

É importante ressaltar que a minimização do MRAP, característica do ajustamento pela ML_{∞} vista na seção 2.1, independe do modelo estocástico desse ajustamento. Porém, a garantia de minimização do MRA no ajustamento pela ML_{∞} é uma propriedade deste ajustamento com pesos iguais, como feito nas aplicações de ABDELMALEK e MALEK (2008) e que foi o caso do exemplo da FIG 1.1.

Já o MRA do ajustamento pela ML_{∞} com pesos diferentes entre si pode não ser o mínimo dentre os dos possíveis métodos de ajustamento. Isso não é incoerente com a teoria, pois a ML_{∞} minimiza sempre o MRAP, mas não necessariamente o MRA para o caso de pesos diferentes entre si. Por exemplo, no ajustamento de observações (ponderadas por diferentes pesos) de uma rede de nivelamento feito por EBONG (1986), o MRA da ML_{∞} foi, inclusive, maior que o do MMQ para a mesma rede. A TAB 4.3 ilustra esse detalhe.

Contudo, o mesmo questionamento surgido na seção 3.1.2 para o caso da ML_1 pode ser feito agora para o ajustamento pela ML_{∞} , em especial para o caso de observações igualmente ponderadas: se sua principal característica é desejável, por que não utilizá-lo como o próprio método de estimação da rede, em substituição ao MMQ? Primeiramente, porque o MMQ também possui características desejáveis no ajustamento, como a garantia de solução única, a mínima variância para os parâmetros estimados e a máxima verossimilhança, considerando a

premissa de ocorrência de somente erros aleatórios normalmente distribuídos nas observações. Além disso, assim como a ML_1 , a ML_∞ possui como limitação a ausência de uma respectiva teoria de propagação de variâncias consolidada na literatura, essencial para uma análise acurada dos resultados do ajustamento de redes geodésicas.

TAB. 4.3: Modelo estocástico da ML_∞ x minimização de MRAP e MRA

Minimiza:	MRAP?	MRA?
ML_∞ com pesos diferentes entre si	Sim	Não necessariamente
ML_∞ com pesos iguais	Sim	Sim

Assim, nessa pesquisa foi investigada a aplicação da ML_∞ na construção do modelo estocástico para posterior ajustamento da rede pelo MMQ, para o qual, dentre outras vantagens citadas naquela seção, tal teoria já é bem conhecida. Algo similar ao que já é feito com o pré-ajuste pela ML_1 para identificação de *outliers*.

Por fim, ressalta-se que, independentemente dos resultados desta pesquisa, a abordagem usual do modelo estocástico não seria preterida. As modelagens foram comparadas por uma questão de análise, mas não foi feita uma escolha entre elas. De fato, pretendeu-se analisar se a modelagem proposta gera resultados mais apropriados para a rede, para verificação da possibilidade de utilização dessa como uma alternativa e de indícios de necessidade de maiores investigações sobre a mesma.

5 PROCEDIMENTOS PROPOSTOS E MÉTODOS DE AVALIAÇÃO

5.1 VISÃO GERAL

A FIG. 5.1 ilustra um possível processo para estimação robusta de redes altimétricas e o posicionamento nele dos procedimentos propostos nessa pesquisa. Ele inicia com o planejamento da rede, seguido da coleta de desníveis no campo (levantamento) e do tratamento dos efeitos sistemáticos de colimação, curvatura da Terra e refração, dentre outros existentes. Esses assuntos, que não estão no escopo dessa pesquisa, foram citados e referenciados na seção 4.2.

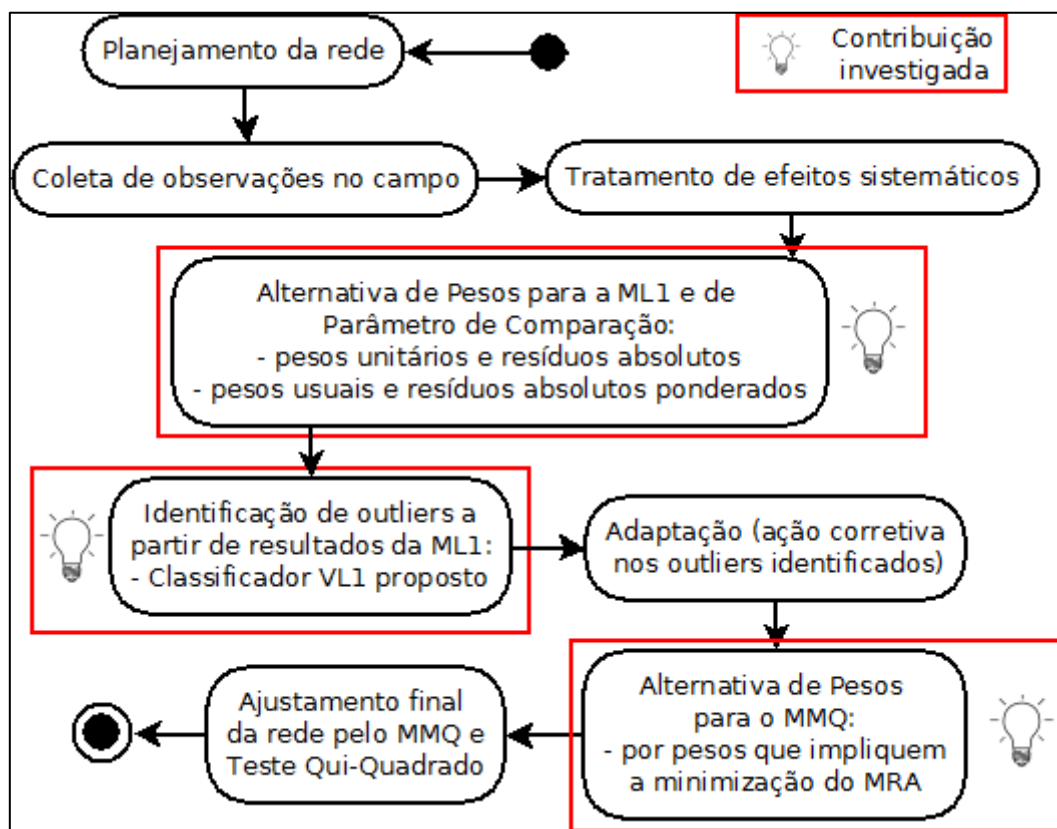


FIG. 5.1: Contribuições dessa pesquisa em um possível processo de estimação robusta.

A seguir, antes da identificação de *outliers* propriamente dita, vem a atividade intitulada *Alternativa de Pesos para a ML1 e de Parâmetro de Comparação*, que será vista na seção 5.2. Nela se desejou investigar abordagens alternativas conjuntas para o modelo estocástico da rede e o parâmetro de comparação na identificação de *outliers* a partir de resultados pela

ML1. As abordagens alternativas testadas foram: pesos unitários com resíduo absoluto como parâmetro de comparação e pesos usuais com o resíduo absoluto ponderado como parâmetro de comparação. Prosseguindo, foi proposto um critério objetivo para a identificação de *outliers* a partir de resultados do ajustamento da rede pela ML1 (Classificador VL1), que será visto na seção 5.3. Vale observar que uma discussão sobre a importância, tanto da *Alternativa de Pesos para a ML1 e de Parâmetro de Comparação*, como da definição de um critério objetivo para a identificação de *outliers* pela ML1, foi feita na seção 3.2.2.

Após a adaptação ou ação corretiva (seção 3.2) sobre os *outliers* identificados (que não é alvo dessa pesquisa), foi investigada, amparada pelo conteúdo visto nas seções 4.3 e 4.4, uma nova abordagem para o modelo estocástico a ser empregado no ajustamento final da rede pelo MMQ: pesos que impliquem a minimização do MRA, obtido a partir de um ajustamento anterior da rede pela ML^∞ com pesos iguais para as observações. Tal atividade, que será vista na seção 5.4, foi chamada de *Alternativa de Pesos para o MMQ*. Em seguida, no processo de estimação robusta apresentado, por fim se considerou a aplicação do Teste *Qui-Quadrado*, para verificação de algum problema no ajustamento.

Vale ressaltar que essa pesquisa está considerando a possibilidade, dentro de um processo de estimação robusta, de se adotar um modelo estocástico na identificação de *outliers* (fase que pode ser considerada um pré-processamento da rede) e outro distinto no ajustamento propriamente dito da rede pelo MMQ.

Quanto às contribuições dessa pesquisa no que se refere à estimação de redes geodésicas, contidas nas hipóteses do trabalho, não foram encontrados na literatura registros sobre:

- a) um critério objetivo para identificação de *outliers* a partir do ajuste pela ML1;
- b) alternativas de modelo estocástico e de parâmetros de comparação para ajuste pela ML1 com a finalidade de detecção e identificação de *outliers*;
- c) alternativas de modelo estocástico para o ajustamento pelo MMQ levando-se em conta um ajuste anterior da rede pela ML^∞ .

Outros desenvolvimentos dessa pesquisa também não foram encontrados na revisão bibliográfica feita. Dentre eles, podem-se ressaltar os seguintes:

- a) apresentação da formulação do ajustamento pela ML^∞ por programação linear para observações com pesos diferentes entre si (seção 2.2.2);
- b) inserção de *outliers* pela *regra dos 3σ para o erro total* (seção 3.1), levando em conta a resistência da geometria da rede a eles pelo MEM (seção 3.3), visando avaliação de métodos de identificação de *outliers*;

c) avaliação da identificação de *outliers* em redes altimétricas a partir de resultados do ajustamento pela ML1 utilizando SMC.

5.2 ALTERNATIVAS DE PESOS PARA A ML1 E DE PARÂMETRO DE COMPARAÇÃO

A *Alternativa de Pesos para a ML1* e a *Alternativa de Parâmetro de Comparação* foram realizadas de maneira conjunta nos experimentos, pois se desejou investigar qual abordagem para o modelo estocástico e com qual parâmetro de comparação, na identificação de *outliers* a partir de resultados do ajustamento da rede pela ML1, apresenta condições mais vantajosas para a execução dessa tarefa. Entende-se por vantajosa a ocorrência do parâmetro de comparação do *outlier* como sendo o maior dentre as observações da rede. Caso isto não ocorra, pode-se considerar que houve mascaramento do *outlier*.

Conforme discussão da seção 3.2.2, a abordagem conjunta usual trata os pesos das observações como proporcionais ao inverso do comprimento das respectivas linhas da rede (pesos usuais) e considera o resíduo absoluto como o parâmetro de comparação para identificação de *outliers*. Com isso, se espera que o resíduo absoluto no ajustamento pela ML1 (com modelo estocástico usual) de um possível *outlier* seja o maior dentre todas as observações da rede, não impossibilitando assim sua identificação. Nessa pesquisa, com aplicação de SMC, isso foi mensurado em redes altimétricas pelo percentual de ocorrências do parâmetro de comparação do *outlier* como sendo o maior dentre as observações da rede (PPCOMR), apresentado na EQ 5.1. Quanto maior o PPCOMR, mais vantajosa a abordagem conjunta avaliada. Similarmente, para cenários de múltiplos *outliers*, o numerador do PPCOMR computou a quantidade de experimentos em que os valores do parâmetro de comparação dos *outliers* foram os maiores da rede.

$$PPCOMR = \frac{\text{experimentos em que o parâmetro de comparação do outlier foi o maior da rede}}{\text{total de experimentos}} \quad (\text{EQ 5.1})$$

O PPCOMR foi avaliado também se adotando pesos unitários com o resíduo absoluto como parâmetro de comparação e pesos usuais com o resíduo absoluto ponderado como parâmetro. Assim, além da abordagem usual, duas alternativas foram testadas, totalizando três comparadas. Obviamente, uma quarta opção, que seria se testar pesos unitários com resíduos absolutos ponderados, não se fez necessária, pois ela é equivalente à abordagem alternativa citada de pesos unitários e resíduos absolutos.

A abordagem conjunta para os pesos e parâmetro de comparação que ocasionou um menor PPCOMR foi considerada a mais apropriada para o ajustamento da rede pela ML1 visando à identificação de *outliers* e foi utilizada pelo Classificador VL1 proposto. A avaliação foi conduzida em cenários de redes nos quais os *outliers* eram conhecidos, para que se pudesse aferir o PPCOMR de cada abordagem conjunta. Dos resultados, foram obtidas conclusões quanto à Hipótese I dessa pesquisa.

5.3 O CLASSIFICADOR VL1 PROPOSTO

Já foi visto que o ajustamento pela ML1 é resistente a *outliers* e, assim, normalmente faz com que os resíduos absolutos desses tendam a ser relativamente maiores. Contudo, a estratégia do Classificador VL1 é determinar um valor de corte (VC), intitulado **VCL1**, para o resíduo absoluto de uma observação no ajustamento da rede pela ML1 com pesos unitários, a partir do qual a mesma passará a ser classificada como *outlier*. Assim, a primeira etapa da aplicação do Classificador é a construção do **VCL1** da rede analisada, caso o mesmo não esteja disponível. A FIG 5.2 ilustra a aplicação completa do Classificador VL1.

Após isso, a rede deve ser ajustada pela ML1, adotando-se modelo estocástico com pesos unitários para as observações. Em seguida, vem a decisão binária sobre a classificação como *outlier* de cada observação da rede, pela comparação de seu respectivo resíduo absoluto $|v_i|$ com o valor de corte **VCL1**. Caso seja maior, considera-se que a *i-ésima* observação é um *outlier* (EQ 5.2).

$$\text{se } |v_i| > VCL1, \text{ a } i\text{-ésima observação da rede é um outlier (teste binário)} \quad (\text{EQ 5.2})$$

Vale esclarecer que a decisão de se utilizar a ML1 com pesos unitários e a confirmação do resíduo absoluto como parâmetro de comparação para o Classificador VL1 está amparada pelos resultados dos experimentos da *Alternativa de Pesos para a ML1 e de Parâmetro de Comparação*, que serão apresentados na seção 6.2. Neles se verificou que com essa abordagem conjunta houve maior PPCOMR.

Para construção do valor de corte **VCL1** do Classificador VL1 (FIG 5.3), cenários aleatórios (com *outlier*) referentes à rede altimétrica analisada devem ser preparados via SMC. A descrição das SMCs utilizadas nessa pesquisa será vista na seção 6.1. Concluídas as simulações, um **VC_k** é estipulado - **k** é o número de ordem do VC. Fazendo **VCL1 = VC_k**, testa-se o comportamento dele para todos os cenários simulados da rede analisada, obtendo-se a respectiva TS (EQ 3.6). Esse processo é iterativo até existir um valor onde a TS seja máxima. Esse é o valor final para **VCL1** a ser empregado para a identificação de *outliers*.

Assim, o processo é repetido quantas vezes se julgue necessário, mudando-se apenas o valor do VC_k estipulado. Obviamente, enquanto houver tendência de crescimento da TS, há indício da necessidade de testes com novos VC_k .

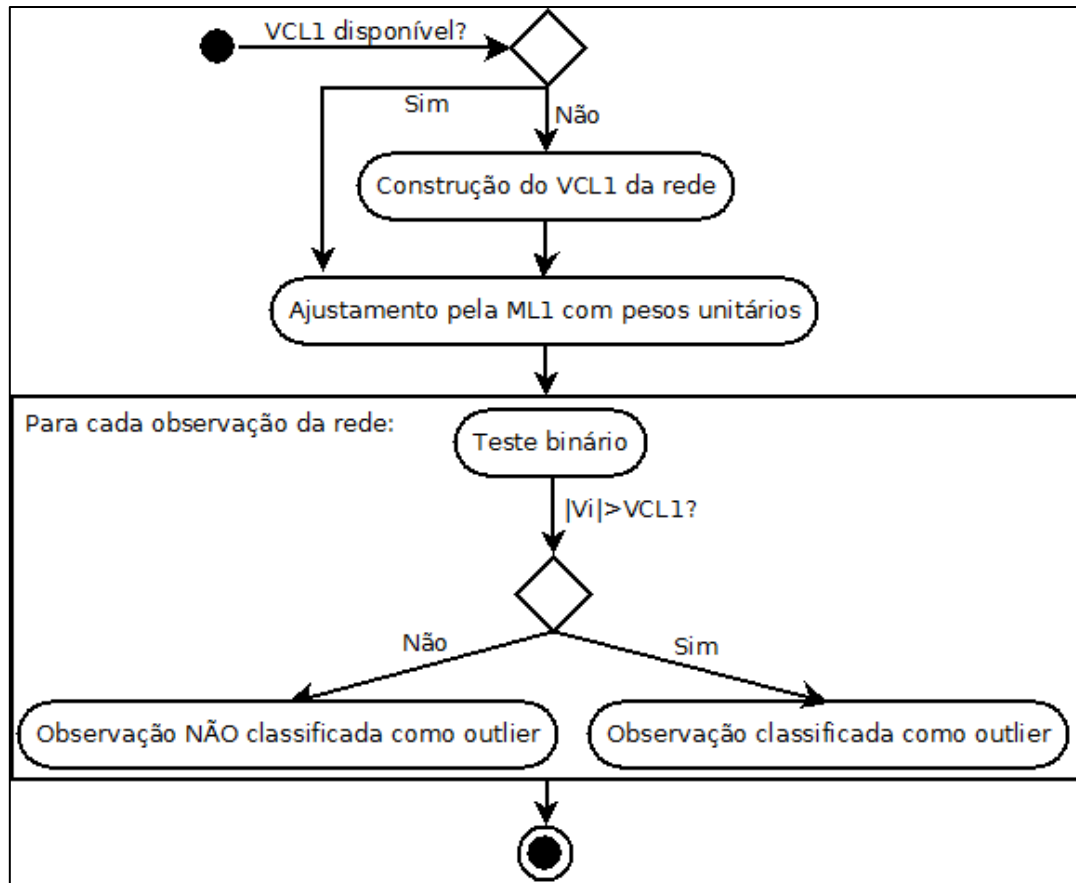


FIG. 5.2: Aplicação do Classificador VL1.

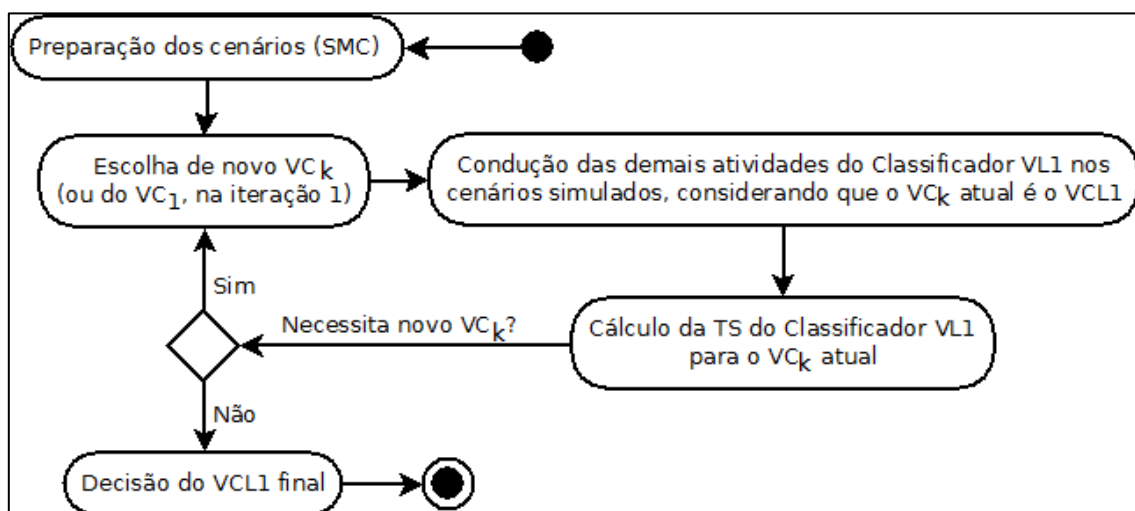


FIG. 5.3: Diagrama de atividades para construção do VCL1.

A aplicação do Classificador VL1 na identificação de *outliers* foi avaliada por comparação com o DS, em relação a cenários de redes nos quais os *outliers* eram conhecidos. A comparação entre os métodos, em ambos os casos, foi feita pelas respectivas TS (EQ 3.6). Dos resultados dessas comparações de desempenho dos classificadores de *outliers* foi possível obter conclusões sobre a Hipótese II dessa pesquisa.

5.3.1 OBTENÇÃO DE UM VC VÁLIDO PARA QUALQUER REDE

A aplicação do Classificador VL1 apresentada até agora considerou apenas o resíduo absoluto da observação no ajustamento pela ML1 para caracterização de *outliers*. Isso se justifica pelos resultados dos experimentos que serão apresentados na seção 6.2, nos quais se demonstrou a vantagem da utilização dele como parâmetro de comparação, em conjunto com modelo estocástico de pesos unitários para as observações.

Entretanto, a obtenção do **VCL1** nesse caso precisa ser repetida para cada rede analisada, pois diferentes redes possuem diferentes geometrias e precisão das observações, o que conduz, obviamente, a diferentes valores de corte para o resíduo dos *outliers*. Assim, seria interessante se investigar fatores que pudessem também ser únicos para toda e qualquer rede altimétrica.

Contudo, outros fatores, apesar de serem tão somente variações do próprio resíduo absoluto, também foram testados, pois se espera que possam ensejar um valor de corte único para identificação de *outliers* em qualquer rede. Tratam-se dos fatores adimensionais **F** (EQ 5.3) e **F'** (EQ 5.4).

$$F_i = \frac{|v_i|}{\text{med}\{|v|\}} \quad (\text{EQ 5.3})$$

$$F'_i = \frac{|v_i|}{\text{mad}\{|v|\}} \quad (\text{EQ 5.4})$$

Nas EQ 5.3 a 5.4, v_i é o resíduo da *i-ésima* observação no ajustamento pela ML1, **med** $\{|v|\}$ a mediana dos valores dos elementos do vetor dos resíduos absolutos $|v|$, desprezando os valores nulos, e **mad** $\{|v|\}$ a mediana dos desvios absolutos dos valores dos elementos de $|v|$ em relação a **med** $\{|v|\}$.

Para esses casos, a aplicação do Classificador VL1 se deu da mesma forma, apenas trocando-se o resíduo absoluto pelo fator estipulado em todas as atividades. Obviamente, essa mudança refletiu sobre o valor do **VCL1**, que, assim, precisou ser reconstruído considerando

a aplicação do Classificador VL1 com cada fator. Para diferenciação, os **VCL1s** obtidos por $|v|$, **F** e **F'** ficam designados como **VCL1_{|v|}**, **VCL1_F** e **VCL1_{F'}**, respectivamente.

A avaliação do Classificador VL1 vista na seção anterior também foi conduzida para os casos de aplicação do Classificador VL1 com os fatores descritos nessa seção. Os **VCL1s** obtidos utilizando os fatores **F** e **F'** foram testados também em rede diferente daquela em que foram construídos, para verificação da possibilidade de aplicação em qualquer rede. Isso tem especial relevância, pois caso positivo indicaria que não é necessário que sejam calculados para cada rede analisada, etapa de maior custo computacional na aplicação do Classificador VL1 (FIG 5.2).

5.4 ALTERNATIVA DE PESOS PARA O MMQ

Quanto aos pesos de observações a serem adotados no ajustamento pelo MMQ após ação corretiva sobre os *outliers* (processo de adaptação), atividade *Alternativa de Pesos para o MMQ*, a abordagem proposta (pesos que impliquem a minimização do MRA) foi comparada com a usual. A FIG 5.4 apresenta o diagrama de atividades para construção do modelo estocástico alternativo proposto.

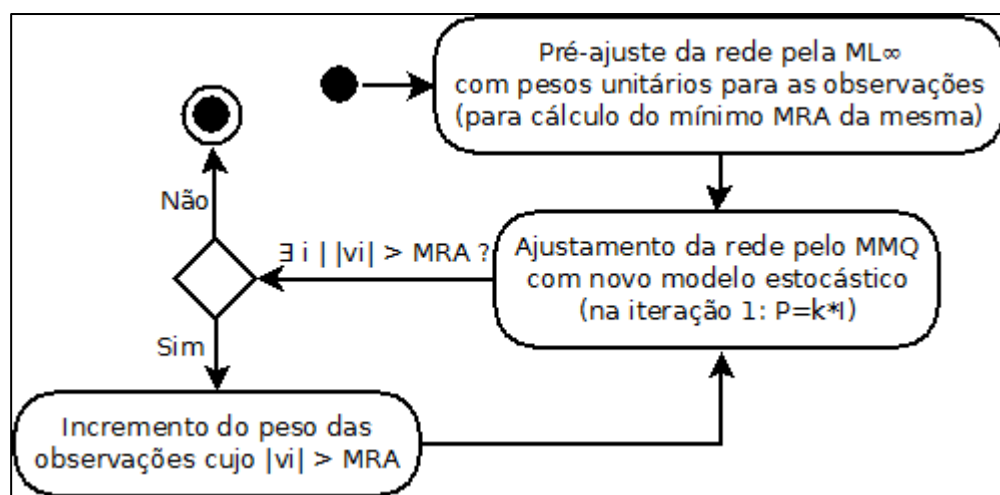


FIG. 5.4: Construção do modelo estocástico alternativo proposto.

Inicialmente, a rede em análise é ajustada pela ML_{∞} com pesos unitários para todas as observações, conforme discussão da seção 4.4, para que se compute o mínimo MRA da mesma. Em seguida, a rede é sucessivamente ajustada pelo MMQ. Na primeira iteração do MMQ, também são atribuídos pesos iguais para as observações, ou seja, a matriz dos pesos **P**

é dada pela matriz identidade **I** de mesmas dimensões multiplicada por um escalar **k**, para o qual, nesse caso, se adotou o valor do inverso do quadrado do MRA. Assim, assume-se que o MRA é uma estimativa inicial para os desvios padrão relativamente consonante com a real qualidade das observações. Nas seguintes, o peso das observações com módulo do resíduo $|v_i|$ superior ao MRA (somado a um nível de tolerância cem vezes menor que a última casa decimal das observações da rede) é incrementado conforme EQ 5.5, sendo **PN_i** o novo peso da observação e **PA_i** o anterior. O procedimento é encerrado quando o módulo de todos os resíduos da rede for menor ou igual ao MRA estipulado. Nesse momento, o modelo estocástico da iteração corrente é o modelo alternativo selecionado, por ter implicado na minimização do MRA no ajustamento pelo MMQ.

$$PN_i = PA_i * \frac{|v_i|}{MRA} \quad (EQ\ 5.5)$$

Para comparação dos resultados, primeiramente foi verificada a ocorrência de algum problema para ambos os ajustamentos pelo MMQ por meio do Teste *Qui-Quadrado*. Uma reprovação no Teste *Qui-Quadrado* sugere a possível ocorrência de algum problema no ajustamento e exigiria uma análise de todo o processo para verificação do mesmo, algo que não está no escopo dessa pesquisa.

Em seguida, seja $|v|$ o vetor do valor absoluto dos resíduos e **DP_x** e **DP_v** o das raízes quadradas dos elementos da diagonal principal das MVCs (que representam as precisões) dos parâmetros estimados e dos resíduos no ajustamento pelo MMQ, respectivamente. Apesar de somente o DP ser relevante para verificação da Hipótese III da pesquisa, pois a mesma se refere à homogeneidade dos resultados, para um maior detalhamento a comparação também foi feita para a média e a máxima magnitude dos elementos dos vetores $|v|$, **DP_x** e **DP_v**, após ajustamento das redes sem *outliers* analisadas pelo MMQ com cada uma das abordagens para o modelo estocástico. Da análise dos resultados foi verificada a possibilidade de confirmação da Hipótese III da presente pesquisa.

6 EXPERIMENTOS E ANÁLISE DE RESULTADOS

Para consecução dos experimentos, observações reais da RAAP e cenários obtidos por SMC a partir de uma rede matemática foram utilizados. A Rede Matemática, base para aplicação da SMC, e a Rede Real são apresentadas na seção 6.1. A razão para a rede que foi base para aplicação da SMC ser chamada de “matemática” será vista na mesma seção. A seção 6.2 detalha o procedimento para simulação de cenários da Rede Matemática sem e com *outliers*. Em seguida, a partir da seção 6.3, os experimentos são descritos, com as respectivas análises de resultados.

Os experimentos foram realizados no software *Octave*. As soluções dos problemas de programação linear foram feitas pelo método *SIMPLEX*, utilizando para tal a função *glpk* do mesmo software com a formulação apresentada no Capítulo 2. A geração de números pseudoaleatórios normalmente e uniformemente distribuídos aplicou as funções *randn* e *rand*, respectivamente. Para a seleção aleatória de observações, a função *randi*, para geração de números inteiros pseudoaleatórios, foi utilizada. O estado inicial das sequências de números pseudoaleatórios foi modificado sempre que necessário para evitar qualquer viés nos resultados. Os códigos *Octave* dos experimentos estão disponibilizados em (SURACI, 2018).

6.1 REDES

6.1.1 REDE MATEMÁTICA – BASE PARA SMC

As observações da rede simulada matematicamente “perfeita” correspondem propositalmente aos desníveis exatos entre as respectivas RRNN. Trata-se de uma rede sem qualquer erro de medida (nem mesmo aleatórios) em suas m observações \mathbf{h}_i ($1 \leq i \leq m$), conforme FIG 6.1 e TAB 6.1, útil para aplicação de SMC, assim como feito por HEKIMOGLU e ERENOGLU (2007), assunto da seção 3.4.1. Por isso foi chamada de Rede Matemática. A altitude do ponto A é considerada fixa e com valor $\mathbf{h}_A = \mathbf{0}$. Assim, trata-se de uma rede com 20 observações e 10 pontos de altitudes a determinar (incógnitas), ou seja, o número de graus de liberdade é 10.

Vale observar que, da análise da geometria da Rede Matemática pelo MEM, conforme discussão da seção 3.3, todas as 20 observações são controláveis contra 1 *outlier* na rede. Para o caso de 2 *outliers*, dos 190 pares possíveis, foram encontrados 24 não controláveis: (1,4),

(1,5), (1,7), (1,8), (4,5), (4,6), (4,8), (5,6), (5,7), (6,7), (6,8), (7,8), (12,16), (12,17), (13,14), (13,20), (14,15), (14,18), (14,20), (15,16), (15,18), (15,19), (16,17), (16,19).

Todos os cenários de rede simulados a partir da Rede Matemática tiveram a mesma geometria da FIG 6.1. Apenas o valor das observações é modificado nas simulações.

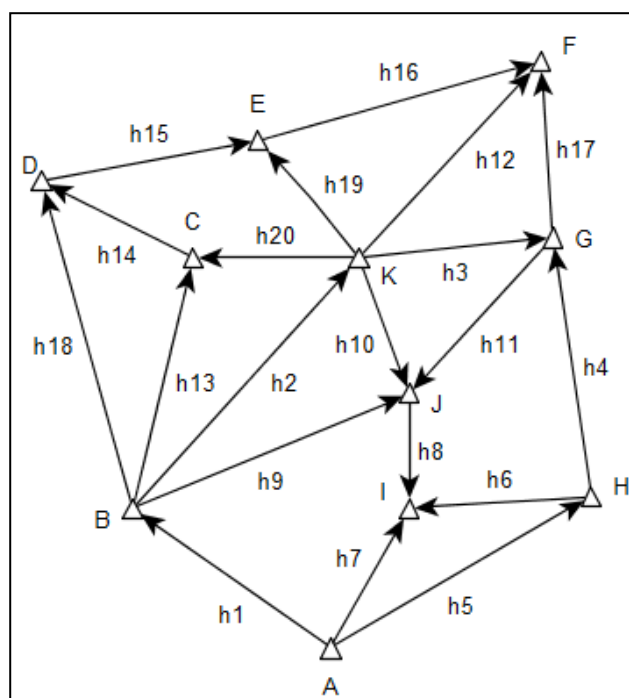


FIG. 6.1: Geometria da Rede Matemática

TAB 6.1: Desníveis exatos – Rede Matemática

h_i	Desnível (mm)	Comp (km)	h_i	Desnível (mm)	Comp (km)
h_1	163854,9	49	h_{11}	110227,2	62
h_2	6446,2	41	h_{12}	155928,2	49
h_3	57037,0	38	h_{13}	52875,0	35
h_4	126209,5	34	h_{14}	62904,2	43
h_5	101128,6	22	h_{15}	3889,5	20
h_6	296885,8	13	h_{16}	42705,7	28
h_7	398014,4	23	h_{17}	98891,2	19
h_8	60449,1	48	h_{18}	115779,2	39
h_9	173710,4	15	h_{19}	113222,5	27
h_{10}	167264,2	24	h_{20}	46428,8	21

6.1.2 REDE REAL - TRECHO DA RAAP

Os dados da RAAP utilizados nos experimentos foram adaptados de (BESERRA, 1952). Tratam-se dos dados do nivelamento geométrico de precisão do território brasileiro que tinha



Fonte: adaptado de BESERRA (1952)

HEKIMOGLU e ERENOGLU (2007) com heterocedasticidade (seção 3.4.1) foi empregada. Foram simulados **100*m** “bons” cenários \mathbf{R}_j ($1 \leq j \leq 100*m$) da rede somente com erros aleatórios. Nas simulações a partir da Rede Matemática, o DP das observações foi dado pela EQ 6.1, sendo \mathbf{K} (em km) o comprimento da respectiva linha.

$$\sigma_i = 1,0(mm) * \sqrt{K_i} \quad (\text{EQ 6.1})$$

As “boas” observações $\mathbf{h}_{i,j}$ ($1 \leq i \leq m, 1 \leq j \leq 100*m$) dos cenários \mathbf{R}_j da rede foram obtidas somando o valor das respectivas \mathbf{h}_i ($1 \leq i \leq m$) e precisão σ_i ($1 \leq i \leq m$), essa última multiplicada por um respectivo número real aleatório $\mathbf{a}_{i,j}$ ($1 \leq i \leq m, 1 \leq j \leq 100*m$) gerado a partir da distribuição normal padrão (EQ 6.2).

$$h_{i,j} = h_i + a_{i,j} * \sigma_i, \quad a_{i,j} \sim N(0,1) \quad (\text{EQ 6.2})$$

Importante ressaltar que, na obtenção das “boas” observações $\mathbf{h}_{i,j}$, apesar da probabilidade de um $\mathbf{a}_{i,j}$ isoladamente ter magnitude maior do que 3 ser “pequena” (0,27%, considerando a normalidade de seu conjunto), pela definição de *outlier* adotada nessa pesquisa, conforme discussão da seção 3.1, caso isso ocorresse sua respectiva observação seria classificada como *outlier*. Assim, como o objetivo aqui foi obter cenários livres de *outliers*, quando isso aconteceu tal valor foi desprezado e um novo $\mathbf{a}_{i,j}$ foi gerado.

6.2.2 SIMULAÇÃO 2 – CENÁRIOS COM *OUTLIERS*

A Simulação 2 corresponde à criação, a partir da Rede Matemática, de inúmeros cenários aleatórios de rede altimétrica com *outliers*. Ela inicia pela realização da Simulação 1. Em seguida, por procedimento também adaptado do caso de heterocedasticidade de HEKIMOGLU e ERENOGLU (2007), para cada um dos cenários \mathbf{R}_j da rede (oriundos da Simulação 1), 100 cenários foram criados pela troca aleatória de uma de suas “boas” observações por um *outlier*, totalizando **10.000*m** ($100 * 100 * m$) cenários \mathbf{R}'_q ($1 \leq q \leq 10.000 * m$) com *outlier*. A FIG 6.3 ilustra a sequência de obtenção dos cenários de rede com 1 *outlier* pela Simulação 2. Esses aparecem em negrito com suas respectivas observações abaixo e entre parêntesis.

Para cada cenário com *outlier*, um *outlier* (EQ 6.3), chamado \mathbf{h}'_q ($1 \leq q \leq 10.000 * m$), substituiu uma das “boas” observações $\mathbf{h}_{i,j}$ da respectiva amostra \mathbf{R}_j da rede. A decisão sobre qual seria a *i-ésima* observação a ser substituída foi feita aleatoriamente, considerando \mathbf{i} um número inteiro ($1 \leq i \leq m$) com distribuição uniforme. Na EQ 6.3, \mathbf{s}'_q ($1 \leq q \leq 10.000 * m$) é uma escolha aleatória entre os valores (-1) ou (1), ou seja, sinal do erro total, e \mathbf{u}'_q

($1 \leq q \leq 10.000 * m$) um número real aleatório entre **LimI** (limite inicial, cujo valor nunca será menor que 3) e **LimF** (limite final) com distribuição uniforme, que representa a magnitude da quantidade de σ_i do erro total e extrapola a *regra dos 3σ para o erro total*, aceita nessa pesquisa para classificação de observações como *outliers* (seção 3.1).

$$h'_q = h_i + s'_{i,j} * u'_q * \sigma_i, \quad u'_q \sim U(\text{LimI}, \text{LimF}) \quad (\text{EQ 6.3})$$

Nessa Simulação 2, os *outliers* foram classificados em quatro grupos: “muito pequenos”, “pequenos”, “médios” e “grandes”, para os quais o erro total está contido nos limites dos intervalos 3σ - 6σ , 6σ - 12σ , 12σ - 25σ e 25σ - 100σ , respectivamente. A TAB 6.2 apresenta os valores de **LimI** e **LimF** (EQ 6.3) para cada grupo.

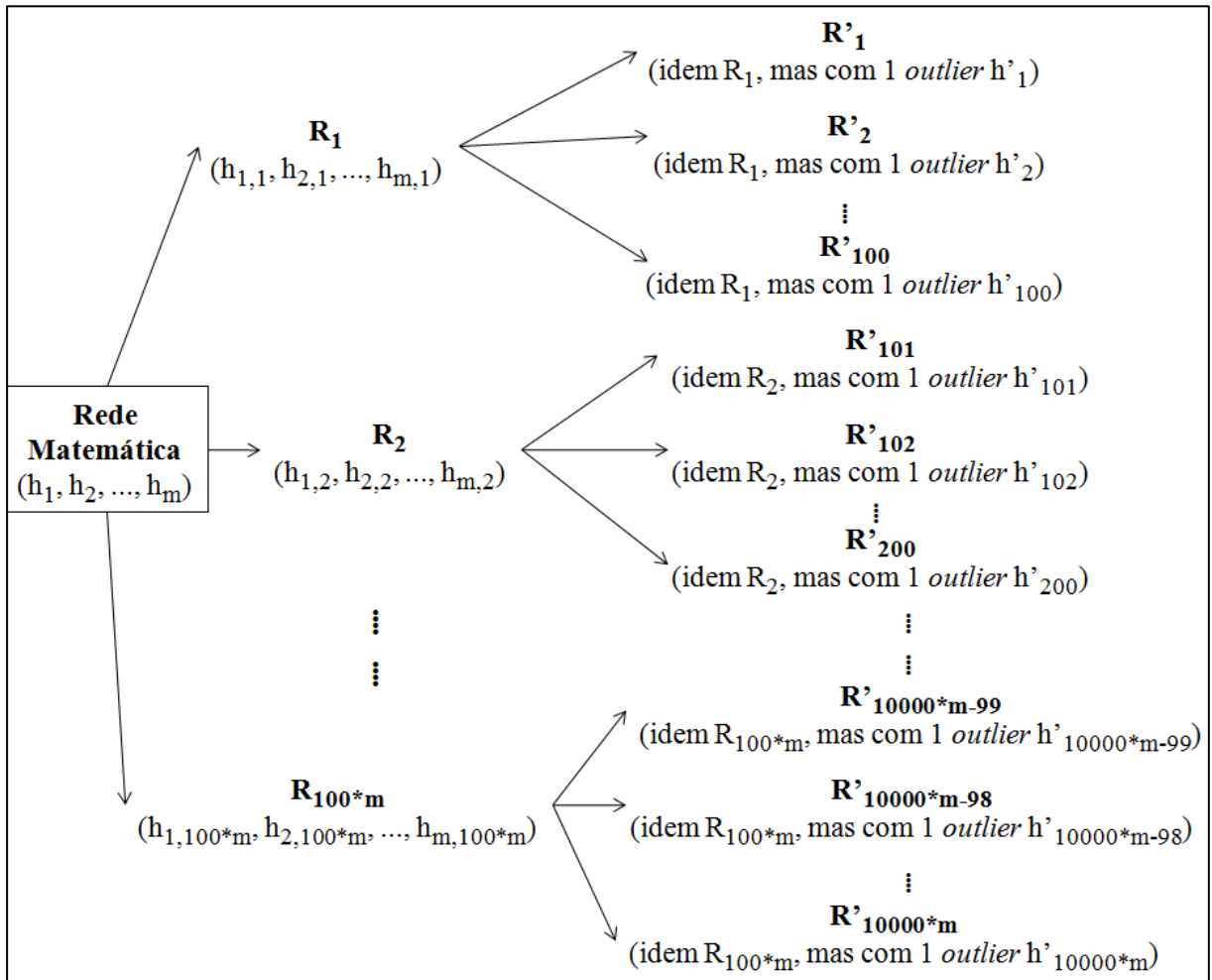


FIG. 6.3: Sequência de obtenção dos cenários de rede com 1 outlier – Simulação 2.

A criação de cenários com múltiplos *outliers* foi similar ao caso de um *outlier* apresentado, apenas com a inserção de mais de um *outlier* em cada cenário R'_q ($1 \leq q \leq 10.000 * m$) com *outlier*.

TAB 6.2: Valores de LimI e LimF para cada tipo de *outlier*

Tipo de <i>outlier</i>	Intervalo	LimI	LimF
muito pequeno	3σ - 6σ	3	6
pequeno	6σ - 12σ	6	12
médio	12σ - 25σ	12	25
grande	25σ - 100σ	25	100

6.3 EXPERIMENTO 1 - ALTERNATIVA DE PESOS PARA A ML1 E DE PARÂMETRO DE COMPARAÇÃO

No Experimento 1, a *Alternativa de Pesos para a ML1 e de Parâmetro de Comparação* (seção 5.2) foi testada em cenários de rede obtidos pela Simulação 2 (seção 6.2.2). A TAB 6.3 apresenta o PPCOMR (EQ 5.1) do *outlier* gerado para diferentes intervalos de magnitude de erro total inserido (implementados pelas variáveis **LimI** e **LimF** da EQ 6.3).

A linha “PIV” corresponde aos resultados dos ajustamentos pela ML1 com pesos pelo inverso da variância das respectivas observações (logo, para o caso da Simulação 2, proporcionais ao inverso do comprimento das linhas de nivelamento – EQ 6.1) e a “PUnit” com pesos unitários para as observações da rede. Já os parâmetros de comparação são indicados por “RA” para o resíduo absoluto e “RAP” para o resíduo absoluto ponderado. O total de cenários de rede simulados, para cada intervalo de magnitude de erro total inserido, foi de 200.000.

TAB 6.3: PPCOMR – cenários com 1 *outlier* – Experimento 1

Erro Total		3-6 σ	6-12 σ	12-25 σ	25-100 σ
PUnit	RA	63,79%	95,01%	99,92%	100,00%
PIV	RA	58,10%	84,32%	89,88%	90,07%
PIV	RAP	59,29%	86,65%	90,05%	90,07%

A TAB 6.4 é similar à TAB 6.3, mas para o caso de 2 *outliers* em cada cenário de rede simulado.

No Experimento 1, o PPCOMR foi sempre maior no ajustamento pela ML1 com pesos unitários. Assim, para as redes analisadas, confirma-se a Hipótese I dessa pesquisa, de que, em redes altimétricas, a adoção de pesos unitários para as observações no ajustamento pela ML1 apresenta condições vantajosas para a identificação de *outliers* em relação aos usuais pesos proporcionais ao inverso do comprimento das linhas de nivelamento.

TAB 6.4: PPCOMR – cenários com 2 *outliers* – Experimento 1

Erro Total		3-6 σ	6-12 σ	12-25 σ	25-100 σ
PUnit	RA	38,02%	79,76%	89,37%	89,66%
PIV	RA	32,06%	62,92%	71,03%	71,37%
PIV	RAP	33,13%	65,64%	71,29%	71,37%

Vale notar ainda que a utilização de pesos usuais com o resíduo absoluto como parâmetro de comparação, abordagem conjunta comum na literatura, apresentou em geral o pior resultado (maior quantidade de mascaramentos) dentre as três avaliadas. Ademais, as três abordagens conjuntas obtiveram aumento do PPCOMR para maiores magnitudes dos *outliers* inseridos e diminuição do PPCOMR com o aumento da quantidade de *outliers*.

6.4 EXPERIMENTOS REFERENTES AO CLASSIFICADOR VL1

6.4.1 EXPERIMENTO 2

O Experimento 2 trata de uma demonstração numérica da estratégia do Classificador VL1, vista na seção 5.3.1, de se estipular um VC para os resíduos absolutos no ajustamento pela ML1 com pesos unitários, visando a identificação de *outliers* na rede. Para viabilizar isso, em um cenário sem *outliers* simulado nos moldes da Simulação 1 (seção 6.2.1), erros grosseiros com magnitude suficientemente grande para gerarem *outliers* foram propositalmente inseridos em duas observações da rede.

A TAB 6.5 apresenta as observações da rede a ser analisada, com destaque para os dois *outliers*, um causado por troca de dígitos no algarismo dos centímetros (h_1) e outro no dos decímetros (h_3). Lembrando que para a Rede Matemática (utilizada como base para a Simulação 1), a geometria da rede e o comprimento das linhas foram vistos nas FIG 6.1 e TAB 6.1. O maior DP de uma observação da rede é de 7,87 mm.

A TAB 6.6 mostra os resíduos absolutos $|v_i|$ das observações da rede no ajustamento pela ML1 com pesos unitários para as observações. Como pode ser visto, os *outliers* apresentaram resíduos absolutos relativamente maiores que os demais, o que é um indicativo de sua ocorrência. Pode-se notar ainda que h_1 , que possui erro grosseiro de menor magnitude que h_3 , já apresentou resíduo absoluto menor que h_3 , ou seja, quanto maior a magnitude do erro do *outlier*, maior tende a ser seu resíduo absoluto.

TAB 6.5: Observações da rede - Experimento 2

h_i	Obs simulada (mm)	Obs final (mm)	h_i	Obs simulada (mm)	Obs final (mm)
h_1	163842,9	163892,9	h_{11}	110222,7	110222,7
h_2	6454,3	6454,3	h_{12}	155935,6	155935,6
h_3	57028,5	57228,5	h_{13}	52872,4	52872,4
h_4	126217,1	126217,1	h_{14}	62894,6	62894,6
h_5	101138,5	101138,5	h_{15}	3887,6	3887,6
h_6	296883,4	296883,4	h_{16}	42706,4	42706,4
h_7	398019,8	398019,8	h_{17}	98890,6	98890,6
h_8	60443,6	60443,6	h_{18}	115772,5	115772,5
h_9	173717,6	173717,6	h_{19}	113219,9	113219,9
h_{10}	167271,4	167271,4	h_{20}	46430,1	46430,1

TAB 6.6: Resíduos absolutos - Experimento 2

h_i	$ v _i$ (mm)	h_i	$ v _i$ (mm)	h_i	$ v _i$ (mm)	h_i	$ v _i$ (mm)
h_1	32,20	h_6	0,00	h_{11}	0,00	h_{16}	9,30
h_2	8,10	h_7	2,10	h_{12}	0,00	h_{17}	3,70
h_3	179,80	h_8	0,00	h_{13}	3,90	h_{18}	0,00
h_4	0,00	h_9	0,00	h_{14}	1,60	h_{19}	0,00
h_5	0,00	h_{10}	0,00	h_{15}	6,00	h_{20}	0,00

Entretanto, a condição de h_2 e h_{16} pode gerar dúvidas, pois seus resíduos absolutos também são relativamente maiores que os das demais observações (excluindo h_1 e h_3), o que poderia conduzir a uma errada conclusão de que são *outliers*. Com isso, surge a necessidade de um critério objetivo, um VC, a partir do qual a observação seja classificada como *outlier*, o que será feito em uma das etapas do Classificador VL1.

Por fim, vale observar que, para o caso analisado, a adoção de qualquer VC entre 9,30 mm e 32,20 mm asseguraria uma correta identificação dos *outliers* da rede.

6.4.2 EXPERIMENTO 3

O Experimento 3 visou comparar o desempenho do Classificador VL1 (seção 5.3) com o do DS na identificação de *outliers*. Para construção do $VCL1_{|v|}$, do $VCL1_F$ e do $VCL1_{F'}$ (seção 5.3.1), foram utilizados cenários de rede com 1 *outlier* obtidos pela Simulação 2 (seção 6.2.2), mas adotando-se $LimI=3$ e $LimF=100$ na EQ 6.3. Merece chamar atenção o fato de que os valores da TAB 6.2 ainda não foram aplicados nesse momento.

A FIG 6.4 ilustra a quantidade de acertos (de um total de 200.000 cenários simulados nessas condições) na identificação de *outliers* para diversos VCs testados, visando estipular o $VCL1_{|v|}$ para a rede analisada. Por ter obtido a maior quantidade de acertos (191.307), o valor

de 29,2 mm foi selecionado. Para cada cenário, foi considerado que houve acerto se o *outlier* foi identificado, sem que nenhuma outra observação também tenha sido assim classificada. Vale notar que com esse $VCL1_{|V|}$, o Classificador VL1 garantiria uma correta identificação de *outliers* no Experimento 2. Os valores obtidos para $VCL1_F$ e $VCL1_{F'}$ foram de 6,9 e 12,9 (adimensionais), respectivamente. O memorial com os VCs testados para definição dos $VCL1_{|V|}$, $VCL1_F$ e $VCL1_{F'}$ do Experimento 3 consta do APÊNDICE 2.

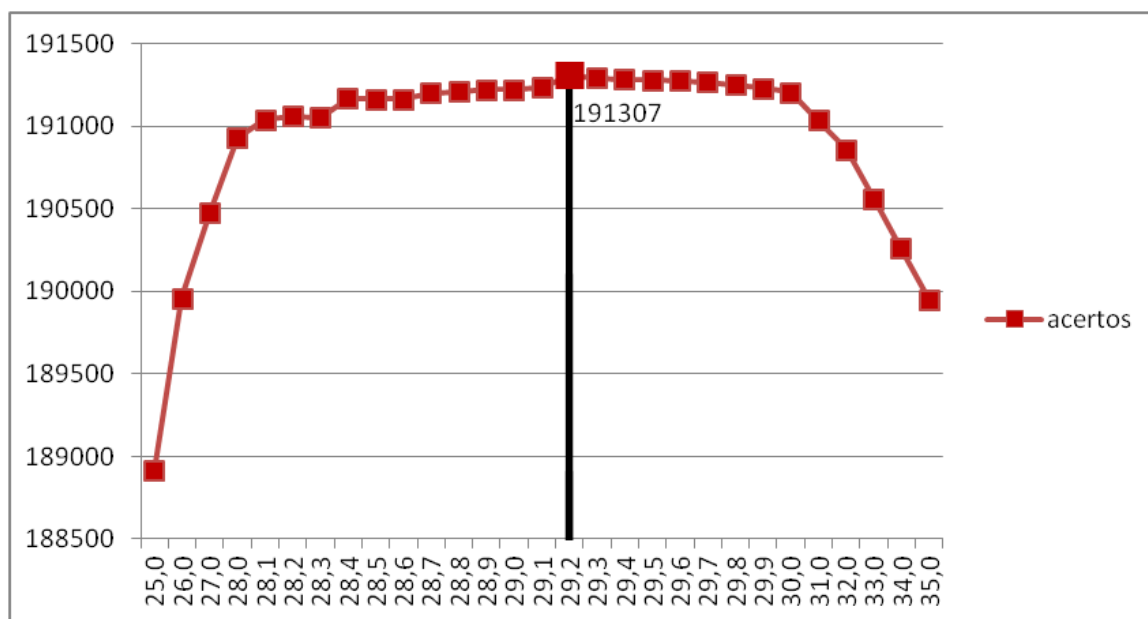


FIG. 6.4: VCs (mm) x quantidade de acertos – escolha do $VCL1_{|V|}$

Em seguida, o Classificador VL1, com o $VCL1_{|V|}$ construído, foi aplicado em cenários de rede, agora com as diferentes magnitudes de *outliers* da TAB 6.2, obtidos via Simulação 2. Para cada faixa de magnitude de *outlier*, foram testados 200.000 cenários simulados. A TAB 6.7 apresenta a TS (EQ 3.6) do Classificador VL1 e do DS nesse contexto. Além disso, o Classificador VL1 também foi testado considerando o $VCL1_F$ e o $VCL1_{F'}$ construídos. Assim, a TAB 6.7 apresenta também as TS do Classificador VL1 com essas 2 outras abordagens para o $VCL1$.

TAB 6.7: TS – 1 *outlier* - Experimento 3

Método	3-6 σ	6-12 σ	12-25 σ	25-100 σ
DS ($\alpha=0,001$)	42,90%	93,92%	98,75%	98,76%
VL1 com $VCL1_{ V }$	18,99%	79,02%	98,93%	99,11%
VL1 com $VCL1_F$	15,02%	59,55%	94,81%	98,22%
VL1 com $VCL1_{F'}$	14,16%	54,64%	92,40%	98,05%

Os resultados do DS apresentados foram obtidos com o recomendado valor de $\alpha=0,001$ (seção 3.2.1) e pesos pelo inverso da variância. Como o Classificador VL1 aplica pesos unitários (motivado pelos resultados do Experimento 1), um teste do DS com pesos unitários foi realizado para comparação também, mas obteve TS inferior a 0,1% para todas as faixas de *outlier*. Assim, essa alternativa foi desprezada na sequência dos experimentos e somente o DS com pesos proporcionais ao inverso da variância passou a ser analisado.

A TAB 6.8 é similar à TAB 6.7, mas para o caso de 2 *outliers* aleatórios inseridos propositalmente em cada cenário de rede simulado. Nesse contexto, foi considerado que houve acerto se ambos *outliers* foram identificados, sem que nenhuma outra observação também tenha sido assim classificada.

TAB 6.8: TS – 2 *outliers* - Experimento 3

Método	3-6 σ	6-12 σ	12-25 σ	25-100 σ
DS ($\alpha=0,001$)	14,32%	76,55%	87,29%	88,88%
VL1 com VCL1 _V	3,35%	56,34%	87,90%	88,32%
VL1 com VCL1 _F	2,17%	24,32%	74,79%	89,06%
VL1 com VCL1 _{F'}	1,92%	19,14%	66,16%	88,81%

As TABs 6.9 e 6.10 são similares à TAB 6.8, mas tratando separadamente de casos de 2 *outliers* controláveis e de 2 não controláveis (considerando a análise da geometria da rede pelo MEM, conforme discussão da seção 3.3), respectivamente. Ambas também apresentam resultados para 200.000 cenários testados. A relação dos 25 pares de observações não controláveis, dentre os 190 pares possíveis, foi vista na seção 6.1.1.

TAB 6.9: TS – 2 *outliers* controláveis - Experimento 3

Método	3-6 σ	6-12 σ	12-25 σ	25-100 σ
DS ($\alpha=0,001$)	16,01%	84,75%	95,99%	97,23%
VL1 com VCL1 _V	3,46%	59,19%	93,48%	93,94%
VL1 com VCL1 _F	2,27%	25,18%	79,04%	94,84%
VL1 com VCL1 _{F'}	2,01%	19,92%	70,00%	94,55%

TAB 6.10: TS – 2 *outliers* não controláveis - Experimento 3

Método	3-6 σ	6-12 σ	12-25 σ	25-100 σ
DS ($\alpha=0,001$)	1,24%	19,39%	27,07%	31,36%
VL1 com VCL1 _V	2,03%	34,56%	49,59%	49,71%
VL1 com VCL1 _F	1,48%	16,62%	44,95%	49,69%
VL1 com VCL1 _{F'}	1,35%	13,03%	40,36%	49,69%

O Experimento 3 confirmou a Hipótese II dessa pesquisa, de que há cenários de redes altimétricas em que a ML1, associado a um critério objetivo adequado para classificação de observações como *outliers*, é mais eficaz que o DS na execução dessa tarefa. Via SMC, o Classificador VL1 apresentou TS superior à do DS em alguns cenários com *outliers*, principalmente para maiores magnitudes desses. Em especial, se ressaltar o resultado do Classificador VL1 em cenários de dois *outliers* não controláveis (TAB 6.10), relativamente bastante superior ao do consagrado DS para *outliers* de magnitude maior que 12σ .

Ademais, pode-se notar que os métodos de identificação de *outliers* analisados tendem a aumentar a eficácia para maiores magnitudes dos *outliers* inseridos e diminuir suas eficácias com o aumento da quantidade de *outliers*.

6.4.3 EXPERIMENTO 4

No Experimento 4, o Classificador VL1 (com os valores obtidos no Experimento 3 para $\mathbf{VCL1_F}$ e $\mathbf{VCL1_{F'}}$) e o DS foram aplicados em 5 cenários obtidos a partir da Rede Real (seção 6.1.2): com 1, 2, 3, 4 ou 5 erros grosseiros propositalmente inseridos nas observações {15}, {15, 27}, {15, 27, 81}, {15, 27, 81, 89} ou {15, 27, 81, 89, 53}, respectivamente. Apesar dos DPs das observações serem desconhecidos, pode-se assegurar que *outliers* foram gerados, pois os erros grosseiros inseridos foram de “grande” magnitude (10 m). Considerando o MEM, a geometria da rede é resistente até mesmo ao cenário de 5 *outliers*, ou seja, em todos os casos os *outliers* gerados são controláveis. A justificativa para se aplicar o Classificador VL1 com $\mathbf{VCL1_F}$ e com $\mathbf{VCL1_{F'}}$, mas não com $\mathbf{VCL1_{|V|}}$, em rede diferente daquela na qual os mesmos foram construídos, foi vista na seção 5.3.4.

A FIG 6.5 compara o tempo médio (em segundos) para o caso de $\mathbf{VCL1}$ disponível (FIG 5.2), de 100 execuções computacionais do DS e do Classificador VL1 (com essas 2 abordagens) na identificação de *outliers* para cada um desses 5 cenários. O computador utilizado tem processador *Intel Core i3* e 4 GB de Memória RAM.

O DS e o Classificador VL1 com $\mathbf{VCL1_{F'}}$ identificaram corretamente os *outliers* nos 5 cenários. Já o Classificador VL1 com $\mathbf{VCL1_F}$, apesar de também identificar corretamente nos 5 cenários, apresentou falsos positivos em 4 deles. Assim, obteve sucesso na identificação apenas em um cenário (o de 5 *outliers*). A TAB 6.11 ilustra os casos em que cada classificador foi bem sucedido na identificação de *outliers*.

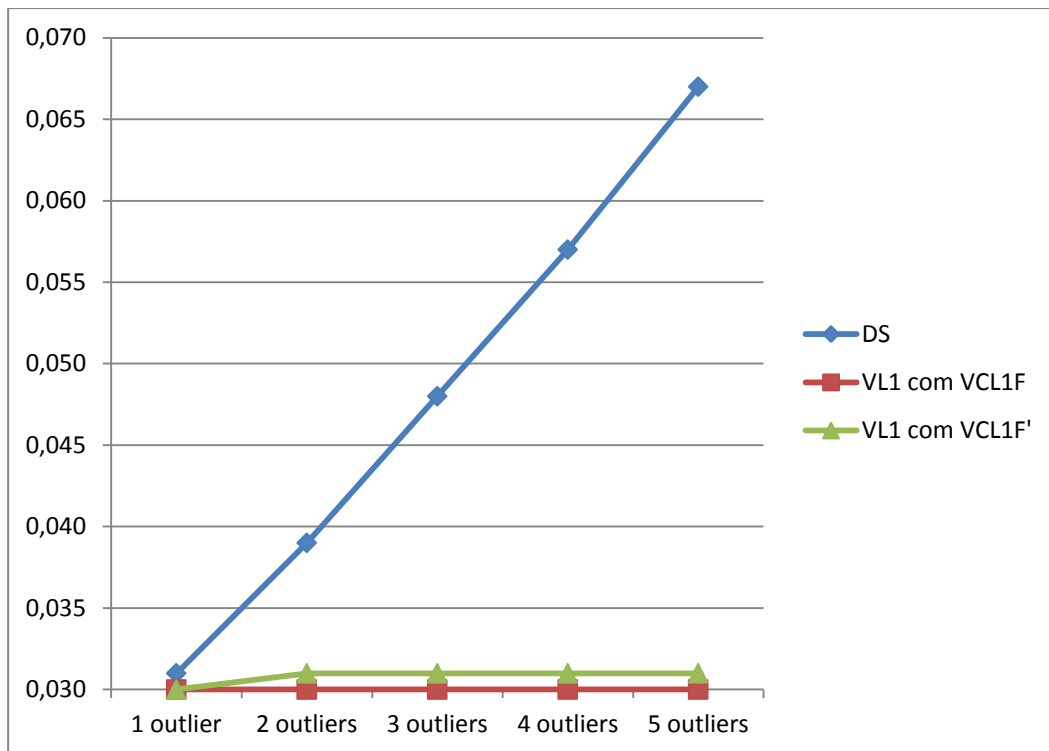


FIG. 6.5: Tempo médio de execução (em segundos) – Experimento 4

TAB 6.11: Sucesso na identificação de *outliers* - Experimento 4

Quantidade de <i>outliers</i>	1	2	3	4	5
DS ($\alpha=0,001$)	X	X	X	X	X
VL1 com VCL1 _F					X
VL1 com VCL1 _{F'}	X	X	X	X	X

O Experimento 4 mostrou que os tempos de execução do Classificador VL1 (com **VCL1** disponível) e do DS são equivalentes para o caso de 1 *outlier*. Porém, o do Classificador VL1 se mantém estável com o aumento da quantidade de *outliers* nas observações, enquanto o do DS apresentou um crescimento de aproximadamente 0,008s a cada *outlier* a mais. Isso representa uma vantagem do Classificador VL1, principalmente em cenários com quantidade relativamente grande de observações, como a RAAP (cerca de 75.000 observações), devido a maior probabilidade de se ter grande quantidade de *outliers*.

Ademais, nesse experimento o Classificador VL1 com **VCL1_{F'}**, mesmo tendo esse sido construído em outra rede (o que é inviável para o **VCL1_{|V|}**), apresentou desempenho superior ao do Classificador com **VCL1_F** de outra rede e idêntico ao do DS. Isso é uma vantagem significativa da aplicação do Classificador VL1 com **VCL1_{F'}** em relação ao Classificador com os demais fatores no caso dessa rede real.

6.4.4 EXPERIMENTO 5

No Experimento 5, o desempenho do Classificador VL1 (seção 5.3), assim como no Experimento 3, foi comparado com o do DS na identificação de *outliers*. Entretanto, para construção do $VCL1_{|v|}$, também utilizando cenários de rede com 1 *outlier* simulados via Simulação 2, adotou-se $LimI=3$ e $LimF=6$ na EQ 6.3, sendo o valor obtido para esse de 17,0 mm (diferente daquele do Experimento 3, para o qual com $LimI=3$ e $LimF=100$, obteve-se $VCL1_{|v|}=29,2$ mm). O memorial com os VCs testados para definição do $VCL1_{|v|}$ do Experimento 5 consta do APÊNDICE 3.

Além disso, o DS foi aplicado também com significâncias $\alpha=0,01$ e $\alpha=0,0001$. A ideia do Experimento foi mostrar que, assim como o DS pode ser aplicado com outros níveis de significância (com implicações para a TS nas diferentes faixas de magnitude de *outliers*), a construção do $VCL1$ também pode ser feita de diferentes formas, com efeitos similares. A TAB 6.12 apresenta os resultados do Experimento 5. Os resultados em negrito destacam, para cada faixa de magnitude de *outlier*, a melhor TS para o Classificador VL1 e para o DS.

TAB 6.12: TS – 1 *outlier* – Experimento 5

Método	3-6 σ	6-12 σ	12-25 σ	25-100 σ
VL1 ($VCL1_{ v }=29,2$ mm)	18,99%	79,02%	98,93%	99,11%
VL1 ($VCL1_{ v }=17,0$ mm)	43,81%	72,22%	74,72%	74,72%
DS ($\alpha=0,001$)	42,90%	93,92%	98,75%	98,76%
DS ($\alpha=0,01$)	54,56%	84,71%	86,66%	86,66%
DS ($\alpha=0,0001$)	26,72%	90,52%	99,91%	99,93%

Os resultados do Experimento 5 revelam que a construção do $VCL1$ pode ser feita de diferentes formas para o resíduo absoluto, com implicações para a TS do Classificador VL1 nas diferentes faixas de magnitude de *outliers*. Por analogia, o mesmo vale o $VCL1$ com os fatores F e F' .

Vale notar que existe relativa homogeneidade na TS para cada método e suas variações nos intervalos de 12-25 σ e de 25-100 σ . Entretanto, tendem a ter uma variação maior da TS entre os intervalos de magnitude dos *outliers* de 6-12 σ e de 12-25 σ , e uma variação ainda mais significativa entre as TS de 3-6 σ e de 6-12 σ . Há indicação ainda que diferentes variações dos métodos sejam melhores para diferentes faixas de *outliers*.

6.5 EXPERIMENTO 6 - ALTERNATIVA DE PESOS PARA O MMQ

No Experimento 6, primeiramente a *Alternativa de Pesos para o MMQ* (seção 5.4) foi testada em 2.000 cenários de rede obtidos pela Simulação 1 (seção 6.2.1). O MRA médio calculado no ajustamento pela ML^∞ dos cenários simulados foi de 6,2162 mm. O modelo estocástico proposto foi obtido, em média, após 34 iterações.

A TAB 6.13 apresenta a quantidade de cenários nos quais os valores máximo, médio e o desvio-padrão dos elementos de $|\mathbf{v}|$, \mathbf{DPv} e \mathbf{DPx} foi menor (considerado mais desejável) para os ajustamentos pelo MMQ com modelo estocástico usual (ME usual), dos pesos pelo inverso da variância das respectivas observações (logo, proporcionais ao inverso dos comprimentos das linhas de nivelamento, para o caso da Simulação 1), e com o modelo estocástico proposto (ME proposto), apresentado na seção 5.4. Os ajustamentos foram reprovados em 105 e 111 dos casos no Teste *Qui-Quadrado* bicaudal ao nível de significância de 5%, para o ME usual e o ME proposto, respectivamente.

Em seguida, a *Alternativa de Pesos para o MMQ* foi testada na Rede Real (seção 6.1.2). A TAB 6.14 apresenta os valores máximo, médio e o desvio-padrão dos elementos de $|\mathbf{v}|$, \mathbf{DPv} e \mathbf{DPx} para os ajustamentos pelo MMQ com o ME usual e com o ME proposto. O MRA calculado no ajustamento pela ML^∞ para a rede foi de 0,1392 m. O modelo estocástico proposto foi obtido após 15 iterações.

TAB 6.13: Quantidade de resultados mais desejáveis – Experimento 6 – Simulação 1

		ME usual	ME proposto
$ \mathbf{v} $	Máximo	0	2000
	Média	1796	204
	DP	35	1965
\mathbf{DPv}	Máximo	165	1836
	Média	1899	101
	DP	153	1848
\mathbf{DPx}	Máximo	1810	190
	Média	1872	128
	DP	241	1759

Ambos os ajustamentos foram reprovados no Teste *Qui-Quadrado* bicaudal ao nível de significância de 5%, porém foram aprovados no Teste *Qui-Quadrado* unicaudal à direita ao nível de significância de 5%. Como nesse experimento os pesos não foram obtidos diretamente a partir dos DPs das observações, pequenas inconsistências de escala no modelo estocástico não são de todo inesperadas. Assim, dado que foram aprovados no Teste *Qui-*

Quadrado unicaudal à direita, é provável que apenas a qualidade das observações possa ter sido subdimensionada, sem ocorrência de problemas de maior relevância nos ajustamentos.

TAB 6.14: Resultados do ajustamento pelo MMQ – Experimento 6 – Rede Real

		ME usual	ME proposto
v	Máximo	0,1896 m	0,1392 m
	Média	0,0219 m	0,0205 m
	DP	0,0281 m	0,0238 m
DPv	Máximo	0,1013 m	0,0406 m
	Média	0,0264 m	0,0313 m
	DP	0,0178 m	0,0042 m
DPx	Máximo	0,1093 m	0,1261 m
	Média	0,0858 m	0,0982 m
	DP	0,0248 m	0,0225 m

Como era de se esperar, o ajustamento pelo MMQ com o ME proposto apresentou MRA (máximo elemento de |v|) inferior em todos os casos (2.000 cenários simulados e rede com dados da RAAP), pois o próprio ME foi construído de forma a implicar isso. Ademais, o DP dos elementos dos vetores |v|, **DPv** e **DPx** nos testes com o ME proposto foi menor, respectivamente, em cerca de 98%, 92% e 88% dos cenários que os equivalentes com o ME usual. Isso mostra que, para a maioria dos cenários analisados, a questão relacionada à Hipótese III dessa pesquisa é verdadeira. O ajustamento pelo MMQ de redes altimétricas livres de *outliers* com modelagem estocástica advinda do MRA gera resultados mais homogêneos que com o modelo usual.

O elemento máximo de **DPv** com ME proposto foi inferior na Rede Real e, em cerca de 92% dos cenários da Simulação 1, também foi menor, o que também pode ser considerado desejável no ajustamento da rede. Em contraponto, os resultados do MMQ com ME usual apresentaram valores médios dos elementos de **DPv** e valores médios e máximos dos elementos de **DPx** relativamente menores em ambos os experimentos. Assim, verifica-se que os elementos máximos de |v| e **DPv** obtiveram comportamento análogo, diferentemente de DPx. Isso sugere que a minimização do MRA pelo ME proposto reflete nos desvios padrão dos próprios resíduos, mas não nos desvios padrão dos parâmetros estimados.

Apenas o valor médio dos elementos de |v| não apresentou resultados equivalentes em ambos os experimentos, tendo sido menor com o ME usual nos cenários simulados, mas menor com o ME proposto para dados da RAAP. Apesar disso ser inconclusivo quanto ao modelo estocástico com resultado mais desejável nessa métrica, de maneira mais geral reforça a importância de que testes sejam sempre realizados também com dados reais.

A desvantagem do ME proposto é que esta ponderação dificilmente reflete a realidade da rede. Neste caso o ajustamento das observações é obtido única e exclusivamente em função dos dados de campo e da configuração geométrica da rede, logo, os resultados obtidos não necessariamente conduzem a uma solução ótima, uma vez que o modelo estocástico foi simplificado. Apesar do ME usual também apresentar essa deficiência, conforme discussão da seção 4.3, espera-se que ele seja uma aproximação realista.

Por fim, diante dos resultados, não se pode indicar qual dos modelos estocásticos devem ser seguidos. Um modelo não exclui o outro. Contudo, os promissores resultados com o ME proposto mostram haver indícios da possibilidade de maiores investigações sobre o mesmo, ou até de formas de combinar os dois modelos estocásticos, visando resultados mais realistas e apropriados.

7 CONSIDERAÇÕES FINAIS

7.1 CONCLUSÕES

A aplicação da norma L_1 (ou Método dos Valores Absolutos) corresponde à minimização da soma dos valores absolutos dos resíduos no ajustamento das observações. Já a da L_∞ (ou Minimização da Norma de *Chebyshev*) corresponde à minimização do máximo resíduo absoluto ponderado do ajustamento. Essa dissertação investigou aplicações da minimização das normas L_1 e L_∞ no contexto da estimação robusta de redes altimétricas, mais especificamente nas etapas de identificação de *outliers* e na definição do modelo estocástico para ajustamento final da rede pelo Método dos Mínimos Quadrados, respectivamente. Os conceitos de *outlier* e estimação robusta foram discutidos, de forma a consubstanciar suas abordagens nessa pesquisa.

Quanto à ML1 na identificação de *outliers*, alternativas de modelo estocástico e de parâmetros de comparação para esse ajuste com essa finalidade foram testadas. Por Simulação de Monte Carlo, em cenários de rede de nivelamento com observações independentes, se verificou que o ajustamento pela ML1 com pesos unitários e com o resíduo absoluto como parâmetro de análise apresentou condições vantajosas para a identificação de *outliers* em relação ao modelo estocástico usual de pesos proporcionais ao inverso do comprimento das linhas de nivelamento, seja esse último considerando resíduo absoluto (única abordagem vista anteriormente na literatura) ou resíduo absoluto ponderado como parâmetro para identificação de *outliers*. Por condições vantajosas entendeu-se o maior percentual de experimentos em que os valores do parâmetro de comparação dos *outliers* foram os maiores da rede.

Ainda sobre a ML1, critérios objetivos para identificação de *outliers* a partir dos resultados desse método de ajustamento foram propostos. Foram estipulados e testadas 3 alternativas para definição de VCs, a saber: 1) para o valor absoluto do resíduo de uma observação no ajustamento pela ML1; 2) para a razão entre esse e a mediana dos resíduos absolutos; 3) e para a razão entre esse e a mediana dos desvios absolutos dos resíduos em relação à mediana dos resíduos absolutos. Todas as alternativas testadas desprezaram os valores nulos no cálculo de medianas dos resíduos absolutos. O procedimento para utilização desses foi intitulado Classificador VL1.

Também por SMC, o Classificador VL1 apresentou TS superior à do DS em alguns cenários com *outliers*, principalmente para maiores magnitudes desses. Em especial, se

ressaltou o resultado do Classificador VL1 em cenários de dois *outliers* não controláveis (considerando o MEM), significativamente superior ao do DS para *outliers* de magnitude maior que 12σ . Principalmente para redes com grande quantidade de observações e baixa controlabilidade função da geometria mais pobre, como a RAAP (cerca de 75.000 observações), isso é particularmente relevante. A implantação de grandes redes altimétricas com geometria que permita grande resistência contra múltiplos *outliers* possui custo elevado, de forma que métodos com maior eficácia em cenários de menor resistência podem ser de grande valia. Os resultados dos experimentos sugeriram ainda, no caso de se calcular o VC em uma rede para aplicação do Classificador VL1 com o VC obtido em rede diferente da primeira, a utilização do **VCL1_F** (opção 3 do parágrafo anterior) é mais vantajosa que a do **VCL1_F** (opção 2 do parágrafo anterior).

Apesar de equivalentes para o caso de 1 *outlier*, para o caso de VC calculado anteriormente, o tempo de execução computacional do Classificador VL1 se mantém estável com o aumento da quantidade de *outliers* nas observações, enquanto o DS apresenta um crescimento contínuo, como era de se esperar, por ser esse último um método iterativo. Isso representa também uma vantagem do Classificador VL1 para cenários com quantidade relativamente grande de observações.

Foi exemplificado também que a construção de VC para o resíduo absoluto no ajustamento pela ML1 pode ser feita de diferentes formas, com implicações para a TS do Classificador nas diferentes faixas de magnitude de *outliers*. Por analogia, o mesmo vale para os outros fatores vistos nessa pesquisa (a razão entre o resíduo absoluto e a mediana dos resíduos absolutos não nulos e para a razão entre esse e a mediana dos desvios absolutos dos resíduos em relação à mediana dos resíduos absolutos não nulos). Isso pode ser útil caso o analista deseje identificar *outliers* com maior sucesso em uma faixa específica de magnitude deles, por exemplo.

Quanto à aplicação da norma L_∞ , um modelo estocástico alternativo para ajustamento final da rede pelo MMQ foi proposto. O modelo aproveitou a característica de minimização do MRA da rede no ajustamento pela ML_∞ com pesos iguais. Utilizando SMC, na maioria dos casos o ajustamento da rede pelo MMQ com o modelo proposto gerou resíduos e precisão desses e dos parâmetros estimados mais homogêneos, com menor desvio padrão, que aqueles com a abordagem usual para o modelo estocástico. Isso pode ser aproveitado para que se evitem redes com vértices relativamente muito menos precisos que outros, especialmente quando o controle é baixo, como na RAAP, que possui apenas uma injunção, por exemplo.

É de todo relevante citar ainda alguns pontos dessa dissertação não vistos na revisão bibliográfica realizada: 1) a apresentação da formulação do ajustamento pela ML^∞ por programação linear para observações com pesos diferentes entre si; 2) a inserção de *outliers* pela *regra dos 3σ para o erro total* e levando em conta a resistência da geometria da rede pelo MEM a eles para avaliação de métodos de identificação de *outliers*; e, 3) a avaliação da identificação de *outliers* em redes altimétricas a partir de resultados do ajustamento pela $ML1$ utilizando SMC.

Diante do exposto, é lícito afirmar que os objetivos propostos foram atingidos, proporcionando alternativas a nível científico e operacional no que se refere ao ajustamento de redes altimétricas por nivelamento direto. Os resultados estimulam a continuidade da pesquisa.

7.2 SUGESTÕES DE TRABALHOS FUTUROS

Como continuidade das investigações dessa dissertação, as seguintes sugestões são apresentadas para trabalhos futuros:

- Novos testes relativos à eficácia na identificação de *outliers* em redes altimétricas do Classificador VL1 proposto com valores de $VCL1_V$, $VCL1_F$ e $VCL1_F'$ obtidos na própria rede a ser analisada;
- Novos testes relativos à eficácia na identificação de *outliers* em redes altimétricas do Classificador VL1 proposto com os valores de $VCL1_F$ e $VCL1_F'$ construídos nessa pesquisa, ou mesmo a definição e teste de outros;
- Verificação da possibilidade de aplicação do Classificador VL1 para redes geodésicas em geral, não somente em redes altimétricas de nivelamento direto;
- Verificação da possibilidade de aplicação do Classificador VL1 para identificação de efeitos sistemáticos em redes geodésicas;
- Desenvolvimento de outros métodos para se estipular um critério objetivo para a identificação de *outliers* a partir de resultados do ajustamento da rede pela $ML1$;
- Proposta de procedimentos iterativos para identificação de *outliers* pela $ML1$. A razão principal da proposta é testar procedimento análogo ao DS, ou seja, aplicação de procedimento sucessivamente e em diferentes iterações, ao invés de executar a identificação de todos os *outliers* em uma única determinação;
- Aplicação de outros métodos de programação linear para solução do ajustamento pela $ML1$, para verificação da possibilidade de resultados vantajosos para a identificação de

outliers (lembrando que a ML_1 , em geral, não apresenta solução única no ajustamento de observações);

- Aplicar processos para descorrelacionamento de observações nos ajustamentos pela ML_1 e ML_∞ ;

- Aprofundamento das aplicações de SMC em redes geodésicas. Em especial, relacionar o grau de confiabilidade dos resultados da SMC com a quantidade de cenários simulados e de observações da rede, entre outros fatores;

- Colocação de uma definição objetiva de *outliers* em redes geodésicas. Dentre outros pontos, deve se discutir, por exemplo, a relevância da *regra dos 3σ para o erro total* aqui apresentada e se o 3σ é de fato o limite mais apropriado nas aplicações geodésicas;

- Comparação da influência de erros no modelo estocástico para a identificação de *outliers* pelo DS e pela ML_1 ;

- Realização de testes do modelo estocástico proposto nessa pesquisa em redes geodésicas em geral, não somente em redes altimétricas de nivelamento direto;

- Análise de outras estatísticas e métricas da qualidade do ajustamento com o modelo estocástico proposto;

- Desenvolvimento de outros modelos estocásticos alternativos para ajustamento de redes altimétricas pelo MMQ a partir de resultados do ajustamento pela ML_∞ ou mesmo pela ML_1 ;

- Desenvolvimento da Teoria de Propagação de Erros para ML_∞ e ML_1 , similarmente à já existente para o MMQ. Com isso, poderia ser testada a possibilidade de o ajustamento pela ML_∞ ou ML_1 vir a substituir o consagrado MMQ; e,

- Busca por outras aplicações para as normas L_1 e L_∞ no contexto da estimação robusta de redes geodésicas, além das abordadas nessa pesquisa.

8 REFERÊNCIAS

- ABDELMALEK, N.; MALEK, W. **Numerical linear approximation in C**. London: CRC Press, 2008, 928p.
- AMIRI-SIMKOOEI, A. Formulation of L1 Norm Minimization in Gauss-Markov Models. **Journal of Surveying Engineering**, p. 37-43, fev., 2003.
- ABNT (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS). **NBR 13133: Execução de levantamento topográfico**. Rio de Janeiro: 1994. 35p.
- BAARDA, W. **A testing procedure for use in geodetic networks**. Publications on Geodesy, New Series, v. 2, n. 5, Delft: 1968. 97p.
- BASELGA, S. Global Optimization Solution of Robust Estimation. **Journal of Surveying Engineering**, p. 123-128, ago., 2007.
- BASELGA, S. Nonexistence of Rigorous Tests for Multiple Outlier Detection in Least-Squares Adjustment. **Journal of Surveying Engineering**, n. 137(3), p. 109–112, aug., 2011.
- BERBER, M. **Robustness analysis of geodetic networks**. Technical Report. University of New Brunswick, Fredericton, Canada, 2006. 138p.
- BESERRA, H. **Ajustamento da rede de nivelamento**. Publicação n. 11 da Série B. Rio de Janeiro: Biblioteca Geográfica Brasileira, Disponível em <https://ww2.ibge.gov.br/home/geociencias/geodesia/artigos/1952_Ajustamento_da_Rede_de_Nivelamento.pdf>, 1952. 33p.
- BUSTOS, O. **Estimação robusta no modelo de posição**. Rio de Janeiro: Instituto de Matemática Pura e Aplicada, 1981, 135p.
- DANTZIG, G. **Linear Programming and Extensions**. Princeton: Princeton University Press, 1963. 656p.
- DURDAG, U. M.; HEKIMOGLU, S.; ERDOGAN, B. Outlier detection by using fault detection and isolation techniques in geodetic networks. **Survey Review**, v. 48, n. 351, p. 400-408, 2016.
- EBONG, M. B. The Chebyshev adjustment of a geodetic levelling network. **Survey Review**, v. 28, n. 220, p. 315-321, 1986.
- GEMAEL, C.; MACHADO, A. M. L.; WANDRESEN, R. **Introdução ao ajustamento de observações: aplicações geodésicas**. 2. ed. Curitiba: Ed. UFPR, 2015. 428p.
- GHILANI, C. **Adjustment Computations: Spatial Data Analysis**. 5. ed. Hoboken: John Wiley & Sons, 2010. 647p.

- GOKALP, E.; GUNGOR, O.; BOZ, Y. Evaluation of Different Outlier Detection Methods for GPS. **Sensors**, n. 8, p. 7344-7358, 2008.
- GOLUB, G.; VAN LOAN, C. **Matrix Computations**. 3. ed. Baltimore e Londres: The Johns Hopkins University Press, 1996. 694p.
- GUI, Q. *et al.* A Bayesian unmasking method for locating multiple gross errors based on posterior probabilities of classification variables. **Journal of Geodesy**, n. 85, p. 191–203, 2011.
- HAWKINS, D. **Identification of Outliers**. New York: Champman and Hall, 1980. 188p.
- HEKIMOGLU, S. *et al.* Detecting Configuration Weaknesses in Geodetic Networks. **Survey Review**, v. 43, n. 323, p. 713-730, out., 2011.
- HEKIMOGLU, S.; ERENOGLU, R. C. Effect of heteroscedasticity and heterogeneousness on outlier detection for geodetic networks. **Journal of Geodesy**, n. 81, p.137–148, 2007.
- HUBER, P.; RONCHETTI, E. **Robust Statistics**. 2. ed. New Jersey: Wiley, 2009, 363p.
- IBGE (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA). **Especificações e Normas Gerais para Levantamentos Geodésicos associados ao Sistema Geodésico Brasileiro**. 2017.
- IAG (INTERNATIONAL ASSOCIATION OF GEODESY). IAG Resolution (No. 1/2015). **For the definition and realization of an International Height Reference System (IHRs)**. Disponível em <https://iag.dgfi.tum.de/fileadmin/IAGdocs/IAG_Resolutions_2015.pdf>, 2015.
- KLEIN, I. **Controle de qualidade de observações geodésicas**. 2011. 316p. Dissertação (Mestrado em Sensoriamento Remoto) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011.
- KLEIN, I. **Proposta de um novo método para o planejamento de redes geodésicas**. 2014. 142p. Tese (Doutorado em Sensoriamento Remoto) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2014.
- KLEIN, I. *et al.* On evaluation of different methods for quality control of correlated observations. **Survey review**, v. 47, n. 340, p. 28–35, 2015.
- KNIGHT, N. L.; WANG, J.; RIZOS, C. Generalised measures of reliability for multiple outliers. **Journal of Geodesy**, n. 84, p. 625–635, 2010.
- KHARAKULEVA, O. **Best L1-Approximation with Splines**. Tese (Mestrado em Ciências). 2013. 58p. Universidade de Oslo, Faculdade de Matemática e Ciências Naturais, 2013.
- LEHMANN, R. On the formulation of the alternative hypothesis for geodetic outlier detection. **Journal of Geodesy**, n. 87, p. 373–386, 2013a.

- LEHMANN, R. The 3σ -rule for outlier detection from the viewpoint of geodetic adjustment. **Journal of Surveying Engineering**, n. 139, v. 4, p. 157–165, 2013b.
- LEHMANN, R.; LOSLER, M. Multiple Outlier Detection: Hypothesis Tests versus Model Selection by Information Criteria. **Journal of Surveying Engineering**, n. 142, v. 4, 2016.
- LEYS, C. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. **Journal of Experimental Social Psychology**, 2013.
- MARSHALL, J.; BETHEL, J. Basic concepts of L1 norm minimization for surveying applications. **Journal of Surveying Engineering**, n. 122, v. 4, p. 168-179, 1996.
- MASON, J. C.; HANDSCOMB, D. C. **Chebyshev polynomials**. Florida: CRC Press, 2003, 335p.
- MICHAELIS. **Dicionário Brasileiro da Língua Portuguesa**. Disponível em <<https://michaelis.uol.com.br/>>, Acesso em: 18 ago. 2018.
- POPE, A. J. **The statistics of residuals and the detection of outliers**. NOAA Technical Report, NOS65, NS1. Rockville: 1976. 33p.
- RAYCHAUDHURI, S. Introduction to Monte Carlo Simulation. **Proceedings of the 2008 Winter Simulation Conference**, p. 91-100, 2008.
- 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.
- SURACI, S. S. **Repositório GitHub de Stefano Sampaio Suraci**. Disponível em: <<https://github.com/stefano998/dissertacaoIME>>, Acesso em: 29 dez. 2018.
- TEUNISSEN, P. **Testing theory: an introduction**. 2. ed. Series on Mathematical Geodesy and Positioning, Delft: Delft Univ. of Technology, 2000. 147p.
- TORGE, W. **Geodesy**. 3. ed. Berlin: de Gruyter, 2011, 416p.
- TRASÁK, P.; STRONER, M. Outlier detection efficiency in the high precision geodetic network adjustment. **Acta Geod Geophys**, n. 49, p. 161–175, 2014.
- VALERO, J.; MORENO, S. Robust estimation in geodetic networks. **Física de la Tierra**, n. 17, p. 7–22, 2005
- WATSON, G. Approximation in normed linear spaces. **Journal of Computational and Applied Mathematics**, n. 121, p. 1-36, 2000.
- WURTEMBERG, I. M. **Ill-posed problems and their applications to climate research**. U.U.D.M. Project Report 2011:16. Uppsala University, Uppsala, Suécia, 2011. 25p.

- XU, Z.; WEINBERGER, K. Q.; CHAPELLE, O. Distance metric learning for kernel machines. **arXiv**, n. 1208.3422, v. 2, 2012.
- YETKIN, M.; BERBER, M. Application of the Sign-Constrained Robust Least-Squares Method to Surveying Networks. **Journal of Surveying Engineering**, n. 139, v. 1, p. 59–65, 2013.
- YETKIN, M.; BERBER, M. Implementation of robust estimation in GPS networks using the Artificial Bee Colony algorithm. **Earth Sci Inform**, n. 7, p. 39–46, 2014.
- YETKIN, M.; INAL, C. L1 norm minimization in GPS networks, **Survey Review**, 43, 323, p. 523–532, 2011.
- YU, H. *et al.* Robust M-estimation using the equivalent weights constructed by removing the influence of an outlier on the residuals. **Survey Review**, 2017.

9 APÊNDICES

9.1 APÊNDICE 1: OBSERVAÇÕES DA REDE REAL

Linha	RN Inicial	RN Final	Desnível observado (m)	Comprimento (km)
1	3L	RN89.39	965,4909	541,46
2	RN89.39	3L	-965,2125	174,26
3	RN89.39	19J	-184,5766	385,84
4	19J	17I	-758,8563	144,17
5	17I	15N	-18,9386	13,00
6	15N	9Y	22,8896	108,10
7	9Y	S.5T	162,2300	248,00
8	S.5T	S.4E	-138,6889	83,43
9	S.4E	P.1K	-36,3277	25,80
10	P.1K	P.1T	175,9286	44,13
11	P.1T	P.2B	-106,2625	71,79
12	P.2B	3L	-82,8555	61,86
13	P.2B	P.1T	106,2602	39,68
14	P.1K	1G	-4,0472	20,33
15	1G	P.2B	73,6905	61,25
16	S.4E	P.4P	-29,9320	59,48
17	P.4P	1G	-10,4241	54,20
18	P.4P	S.5N	-12,4700	73,95
19	S.5N	S.5T	181,0665	26,20
20	S.5N	8P	-5,0135	67,08
21	8P	9Y	23,8215	111,15
22	8P	15N	0,9243	146,63
23	19J	21A	134,8814	114,21
24	21A	21B	-17,4816	2,32
25	21B	17I	-876,2539	125,56
26	21A	MMItapeva	-233,2686	449,34
27	MMItapeva	Apiaí	243,1308	88,97
28	Apiaí	21B	-27,2005	170,71
29	Apiaí	4IGG	-220,1132	97,77
30	4IGG	MMItapeva	-23,0415	71,40
31	4IGG	KM98,5	75,4206	130,45
32	KM98,5	Itu	-186,3108	69,74
33	Itu	RN21	91,6736	192,64
34	RN21	MMItapeva	-3,5715	187,75
35	Itu	RN9	36,9309	116,91
36	RN9	Ipiranga	141,9885	207,94

Linha	RN Inicial	RN Final	Desnível observado (m)	Comprimento (km)
37	KM98,5	Ipiranga	-7,4500	104,57
38	Ipiranga	KM98,5	7,4489	343,76
39	RN9	38A	-86,6274	119,70
40	38A	30IGG	-56,6636	179,97
41	30IGG	RN21	197,9122	166,83
42	38A	40Y	200,2670	17,69
43	40Y	42F	-209,1825	83,94
44	42F	Olímpia	-34,4465	149,01
45	Olímpia	30IGG	-13,2047	142,94
46	Olímpia	45O	-12,1858	152,97
47	47O	47B	314,2420	106,10
48	47B	49G	198,3960	144,99
49	49G	42F	-466,0313	102,18
50	45O	52Y	244,7099	170,45
51	52Y	62Z	203,2885	116,89
52	62Z	67S	-66,3206	224,55
53	67S	69M	317,4607	121,30
54	69M	47B	-384,9907	250,25
55	52Y	62Z	203,3870	810,85
56	81J	69M	469,5058	80,56
57	81J	85U	87,0408	248,13
58	85U	67S	65,0532	733,15
59	49G	75G	-122,2306	154,27
60	75G	78O	135,8454	231,57
61	78O	81J	-296,4706	162,15
62	85U	87Q	90,3870	97,77
63	87Q	90K	100,9558	201,41
64	90K	70O	18,0729	132,89
65	90K	93J	-87,0128	98,42
66	93J	94S	4,1540	100,45
67	94S	95Z	-34,5515	91,51
68	95Z	75G	-0,1483	111,09
69	95Z	40Y	-134,7893	257,50
70	94S	RJ.18V	-426,1270	217,55
71	RJ.18V	100B	38,1722	61,75
72	100B	RN9	104,8925	328,83
73	Ipiranga	100B	-246,7157	232,82
74	RJ.18V	RJ.15E	-153,2391	248,80
75	RJ.15E	115N	343,9193	46,46
76	115N	121J	480,8950	94,40
77	121J	93J	-249,5430	189,77
78	121J	106I	-352,5339	216,21
79	106I	87Q	89,0700	16,06

Linha	RN Inicial	RN Final	Desnível observado (m)	Comprimento (km)
80	106I	126R	-405,2769	257,15
81	126R	119K	-63,4986	122,26
82	119K	115N	340,3589	126,99
83	119K	131T	-137,8070	158,63
84	131T	134T	569,4108	123,99
85	134T	126R	-368,0582	150,63
86	134T	148L	-165,7041	21,08
87	176Z	148L	586,3355	247,84
88	176Z	RJ.6D	94,7616	70,63
89	RJ.6D	RJ.7P	0,8766	99,61
90	131T	RJ.7P	-86,8668	78,20
91	148L	176Z	-586,4039	545,71
92	RJ.7P	RJ.10K	373,2157	187,89
93	RJ.10K	RJ.1V	-459,8809	143,85
94	RJ.1V	RJ.1O	-15,3156	24,01
95	RJ.1O	RJ.6D	101,1063	306,49
96	RJ.10K	RJ.15E	-152,0357	196,96
97	RJ.1Z	RJ.18V	461,3395	302,18
98	RJ.1V	RJ.1Z	-0,2796	7,39
99	RJ.1Z	RJ.1H	0,5831	15,60
100	RJ.1H	RJ.1J	49,0682	8,94
101	RJ.1J	RJ.1O	-64,7117	23,22
102	RJ.1H	RJ.1E	34,7145	3,17
103	RJ.1E	RJ.1F	-31,8426	0,99
104	RJ.1F	RJ.1J	46,1939	7,35
105	RJ.1F	RJ.1E	31,8381	25,53

9.2 APÊNDICE 2: MEMORIAL DOS PRINCIPAIS VCs TESTADOS – EXPERIMENTO 3

VCL1 _V		VCL1 _F		VCL1 _{F'}	
VC (mm)	acertos	VC	acertos	VC	acertos
25,0	188912	4,0	166836	10,0	182326
26,0	189959	5,0	180652	11,0	183646
27,0	190478	6,0	185081	12,0	184229
28,0	190929	6,1	185258	12,3	184321
28,1	191042	6,2	185445	12,4	184343
28,2	191060	6,3	185525	12,5	184381
28,3	191051	6,4	185593	12,6	184452
28,4	191168	6,5	185643	12,7	184494
28,5	191165	6,6	185712	12,8	184506
28,6	191165	6,7	185820	12,9	184508
28,7	191202	6,8	185905	13,0	184501
28,8	191212	6,9	185941	13,1	184482
28,9	191223	7,0	185890	13,2	184458
29,0	191222	7,1	185813	13,3	184446
29,1	191233	7,2	185797	13,4	184379
29,2	191307	7,3	185694	14,0	184103
29,3	191292	7,4	185565	15,0	183435
29,4	191284	7,5	185466	16,0	182646
29,5	191279	8,0	185070	xxx	xxx
29,6	191278	9,0	183643	xxx	xxx
29,7	191267	10,0	181879	xxx	xxx
29,8	191252	xxx	xxx	xxx	xxx
29,9	191230	xxx	xxx	xxx	xxx
30,0	191205	xxx	xxx	xxx	xxx
31,0	191038	xxx	xxx	xxx	xxx
32,0	190859	xxx	xxx	xxx	xxx
33,0	190563	xxx	xxx	xxx	xxx
34,0	190261	xxx	xxx	xxx	xxx
35,0	189947	xxx	xxx	xxx	xxx

9.3 APÊNDICE 3: MEMORIAL DOS PRINCIPAIS VCS TESTADOS – EXPERIMENTO 5

VCL1 _v	
VC (mm)	acertos
16,0	89238
16,1	89737
16,2	89887
16,3	89987
16,4	89891
16,5	89971
16,6	90035
16,7	90037
16,8	90313
16,9	90260
17,0	90486
17,1	90398
17,2	90322
17,3	90331
17,4	90235
17,5	90029
17,6	90004
17,7	90018
17,8	89836
17,9	89472
18,0	89358