

## CONFIABILIDADE INTERNA E EXTERNA DURANTE A ETAPA DE PLANEJAMENTO DO LEVANTAMENTO: EXEMPLO EM UMA REDE DE NIVELAMENTO

MARCELO TOMIO MATSUOKA<sup>1</sup>  
JOÃO FRANCISCO GALERA MONICO<sup>2</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Sul – UFRGS  
Instituto de Geociências  
Departamento de Geodésia, Porto Alegre – RS

<sup>2</sup>Universidade Estadual Paulista - Unesp  
Faculdade de Ciências e Tecnologia - FCT  
Departamento de Cartografia, Presidente Prudente - SP  
tomio.matsuoka@gmail.com; galera@fct.unesp.br

**RESUMO** – No contexto do Ajustamento de Observações, um tema ainda pouco explorado pela comunidade geodésica brasileira é a confiabilidade interna e externa. O objetivo deste artigo é apresentar como se pode obter um valor aproximado de confiabilidade interna e externa ainda na etapa de planejamento de um levantamento de campo, devendo ser mais um dos aspectos a ser considerado nesta fase do trabalho. É apresentada uma revisão sobre os testes estatísticos normalmente utilizados no Ajustamento de Observações, o Teste Global e o Data Snooping, e sobre Confiabilidade interna e externa. Um exemplo de aplicação do cálculo da confiabilidade interna e externa em uma rede de nivelamento na etapa de planejamento é apresentado e discutido.

**ABSTRACT** – In the context of Adjustment, reliability (internal and external) is a subject not very well explored by the Brazilian Geodetic Community. The goal of this paper is to present how the internal and external reliability can be computed in the stage of field surveying planning. It is presented a revision about the following subjects: statistical tests normally used in the Adjustment (Global Test and Data Snooping); and Internal and External Reliability. At the end, an example of application of the computation of the internal and external reliability in a leveling network in the stage of field surveying planning is presented and discussed.

### 1 INTRODUÇÃO

Em Geodésia, tal como em muitas outras ciências, o ajustamento de observações tem por objetivo proporcionar uma solução única para parâmetros em problemas onde o número de observações é redundante e o sistema de equações a solucionar inconsistente, bem como uma estimativa da qualidade da solução. A inconsistência do sistema de equações é devida às flutuações probabilísticas das observações e faz com que um determinado subconjunto de dados proporcione valores diferentes de outro subconjunto. A solução única nesses problemas é dada, em geral, pelo Método dos Mínimos Quadrados (MMQ) (GEMAEL, 1994).

As observações apresentam flutuações probabilísticas, pois, quando se repete a medida de uma grandeza, os vários valores não são idênticos, mas dispersos numa certa região ou intervalo. Essas flutuações, tradicionalmente, são classificadas como erros aleatórios. Além dos erros aleatórios, as observações podem também estar contaminadas pelos erros

sistemáticos e grosseiros (GEMAEL, 1994). Do ponto de vista estatístico, observações com erros grosseiros não podem ser consideradas como pertencentes à amostra, não podendo ser usadas com as outras observações. Desta forma, é importante dispor de técnicas para a detecção e identificação de erros grosseiros nas observações (TEUNISSEN, 2000).

Existem duas razões principais para o uso de observações redundantes no ajustamento (TEUNISSEN, 2000). A primeira visa melhorar a acurácia dos resultados estimados. A segunda visa detectar a possível presença de erros (não aleatórios) nos modelos e/ou nas observações.

Dessa forma, a redundância de dados possibilita a aplicação de controle de qualidade nos resultados do ajustamento. Na prática, o controle de qualidade já se inicia na fase de planejamento. Os fatores considerados no controle de qualidade dividem-se em: economia, precisão e confiabilidade. A economia é um fator importante num projeto, pois expressa o custo total, envolvem operações de planejamento, execução, coleta de dados, processamento, etc. A precisão manifesta a

dispersão dos resultados, sendo expressa pela Matriz Variância-Covariância (MVC) dos parâmetros. A confiabilidade descreve a capacidade das observações em detectar erros nos modelos, bem como nas próprias observações, com certo nível de probabilidade. Trata-se nesse caso da confiabilidade interna. A influência dos erros não detectados nos parâmetros permite que se avalie a confiabilidade externa.

Na etapa do controle de qualidade do ajustamento, os testes estatísticos global e o data *snooping* são amplamente utilizados e conhecidos pela comunidade geodésica. Porém, poucos fazem a análise da confiabilidade interna e externa antes ou após o Ajustamento.

Um fato que chama atenção é que a confiabilidade interna e externa pode ser analisada antes mesmo de ser executado o levantamento de campo, ou seja, na etapa de planejamento. Neste sentido, a confiabilidade pode auxiliar também na etapa de planejamento para que se definam estratégias que possam garantir a qualidade final requerida.

O objetivo do presente artigo é apresentar uma revisão dos conceitos de confiabilidade interna e externa para fins de Geodésia e apresentar um exemplo de aplicação em uma rede de nivelamento da análise da confiabilidade interna e externa durante a fase de planejamento do levantamento de campo. Este artigo é uma continuação do estudo apresentado por Monico et al. (2006), no qual abordou a análise da confiabilidade interna e externa da rede após o Ajustamento ter sido realizado.

## 2 TESTANDO A QUALIDADE DO AJUSTAMENTO

Os testes estatísticos normalmente aplicados no controle de qualidade são o Teste Global e o Data Snooping (TEUNISSEN, 2000), as quais são apresentadas a seguir.

### 2.1 Teste estatístico para detecção de erros (Teste Global do Ajustamento)

O teste estatístico para detecção de erros é conhecido como Teste Global do Ajustamento (TEUNISSEN, 1998) ou, também, por teste Qui-quadrado (GEMAEL, 1994).

O fato de se escolher um valor qualquer para o fator de variância *a priori* ( $\sigma_0^2$ ) não acarreta nenhum efeito no resultado do ajustamento. Essa escolha é feita antes do ajustamento para o cálculo da matriz dos pesos das observações (GEMAEL, 1994).

Após o ajustamento, pode-se estimar um valor para o fator de variância em função dos resíduos, que é chamado de fator de variância *a posteriori* ( $\hat{\sigma}_0^2$ ).

A comparação entre os valores de  $\sigma_0^2$  e  $\hat{\sigma}_0^2$  pode ser um indicador da qualidade do ajustamento. Primeiramente, se houver discrepância entre  $\sigma_0^2$  e  $\hat{\sigma}_0^2$

( $\sigma_0^2 \neq \hat{\sigma}_0^2$ ) aplica-se um teste de hipótese baseado na distribuição  $\chi^2$  para constatar se a discrepância é significativa a um certo nível de confiança (GEMAEL, 1994). Uma resposta afirmativa indica que existem problemas no ajustamento.

Desta forma, o teste estatístico  $\chi^2$  testa a hipótese nula:

$$H_0: \sigma_0^2 = \hat{\sigma}_0^2 \text{ contra } H_a: \sigma_0^2 \neq \hat{\sigma}_0^2 \quad (1)$$

As hipóteses formuladas neste teste descrevem que, na hipótese nula ( $H_0$ ), os fatores de variância *a posteriori* e *a priori* são estatisticamente iguais e na hipótese alternativa ( $H_a$ ) o contrário. Para testar se a hipótese nula é rejeitada ou não, a um nível de significância  $\alpha$ , compara-se o valor calculado por:

$$T = (\sigma_0^2 / \hat{\sigma}_0^2) (n - u) = \frac{V^T P V}{\sigma_0^2} \quad (2)$$

com o valor teórico da distribuição  $\chi^2_{(n-u, \alpha)}$ . Onde:

$n$  : é o número de observações;

$u$  : é o número de parâmetros;

$V$  : é o vetor ( $n \times 1$ ) dos resíduos das observações e

$P$  : é a matriz peso das observações ( $P = \sigma_0^{-2} \sum_{l,b}^{-1}$ , onde  $\sum_{l,b}^{-1}$  é a MVC das observações).

A hipótese nula “não” é rejeitada a um nível de significância  $\alpha$ , se:

$$T < \chi^2_{(n-u, \alpha)} \quad (3)$$

Caso contrário significa que há problemas no ajustamento. Assim, uma análise deve ser feita para verificar as possíveis razões ou circunstâncias que levaram à falha do teste, dentre as quais se deve verificar se o modelo funcional é adequado para a realidade física; se o modelo estocástico é apropriado; se não existem erros grosseiros e sistemáticos; se o sistema não é mal condicionado e se não existe erro de cálculo no processo de ajustamento. Neste trabalho o foco é com relação a problemas devido a erros grosseiros nas observações.

Um detalhe importante é que o teste apresentado detecta o problema, mas não explicita qual. Neste sentido, no próximo tópico será apresentado um teste para a identificação da observação com erro grosseiro (outlier).

### 2.2 Teste estatístico para a identificação das observações com erro grosseiro

Para abordar este tópico, assume-se que no caso em que o Teste Global do Ajustamento falhar, qualquer especificação inadequada relacionada com o modelo funcional, bem como erros nos cálculos, possam ser identificadas. Este artigo se detém com a revisão sobre a técnica de identificação de erros nas observações (*outliers*).

No processo de identificação de erros é comum comparecer o termo “outlier” como sinônimo de erro grosseiro. Um “outlier” é definido como um resíduo que contradiz algumas propriedades estatísticas. Isto possibilita definir uma estratégia de teste estatístico para a detecção de outliers, dependendo, portanto, do nível de risco e distribuição de probabilidade assumidos, bem como do procedimento do teste. Independente da diferença entre a definição de erro grosseiro e outlier, assume-se que os outliers identificados são causados por erros grosseiros. As estratégias de identificação de outliers têm suas raízes nos trabalhos do Professor Baarda da Universidade Técnica de Delft na Holanda (BAARDA, 1968). A técnica desenvolvida por Baarda (1968) é denominada de Data Snooping e será revisada a seguir.

**Método de Baarda: Data Snooping**

Neste método utiliza-se a convenção de que somente um erro grosseiro (outlier) está presente no conjunto de observações. Desta forma, para identificar a i-ésima observação com a presença de um outlier, o data snooping baseia-se no cálculo da seguinte expressão (TEUNISSEN, 2000):

$$w_{lb(i)} = \frac{c_{lb(i)}^T P V}{\sqrt{c_{lb(i)}^T P \sum_v P c_{lb(i)}}}, \tag{4}$$

onde P é a matriz dos pesos das observações,  $\sum_v$  é a matriz variância-covariância dos resíduos, e o vetor  $c_{lb(i)}$  é um vetor n-dimensional contendo um valor unitário na linha da i-ésima observação a ser testada e zero nas demais, ou seja:

$$c_{lb(i)}^T = [0 \dots 0 \dots \underset{1}{1} \dots 0 \dots 0] \tag{5}$$

Em muitas aplicações práticas, a matriz variância-covariância das observações é diagonal, ou seja, não se considera ou não existe correlação entre as observações. Nestes casos, a equação (4) se reduz para (TEUNISSEN, 2000):

$$w_{lb(i)} = \frac{v_i}{\sigma_{v_i}}, \tag{6}$$

onde  $v_i$  é o resíduo da i-ésima observação a ser testada e  $\sigma_{v_i}$  é o respectivo desvio-padrão desse resíduo.

A estatística  $W_{lb(i)}$  tem como distribuição a raiz quadrada da distribuição F, com graus de liberdade igual a 1 e  $\infty$  (infinito). A raiz quadrada da distribuição F tem

distribuição normal padrão, isto é,  $\sqrt{F_{\alpha_0(1,\infty)}} = N_{\alpha_0/2}(0,1)$ . Então (TEUNISSEN, 2000):

$$w_{lb(i)} \sim N_{\alpha_0/2}(0,1) \tag{7}$$

As hipóteses formuladas descrevem que, na hipótese nula a i-ésima observação não contém outlier ( $\nabla_{lb(i)}$ ) e na hipótese alternativa o oposto. Dessa forma, testa-se:

$$H_0: \nabla_{lb(i)} = 0 \quad \text{contra} \quad H_a: \nabla_{lb(i)} \neq 0 \tag{8}$$

A hipótese básica não é rejeitada a um determinado nível de significância  $\alpha_0$  ( $\alpha = 2 n \alpha_0$ , desigualdade de Bonferroni (CHEN et al., 1987)) se (TEUNISSEN, 2000) :

$$|W_{lb(i)}| \leq N_{\alpha_0/2}(0,1). \tag{9}$$

Caso contrário, a i-ésima observação é suspeita de possuir um outlier. Como se trata de um teste que verifica somente uma observação por vez e dependendo da magnitude do erro (outlier -  $\nabla_{lb(i)}$ ), após calcular o valor  $w_{lb(i)}$  para todas as observações, pode acontecer de mais do que um ou até todos excederem o valor crítico  $N_{\alpha_0/2}(0,1)$ . Neste caso, a observação suspeita será aquela que possuir o maior valor para  $w_{lb(i)}$ . Logo, de uma forma geral, pode-se definir que a j-ésima observação será suspeita de possuir um outlier se (TEUNISSEN, 2000):

$$|W_{lb(j)}| > N_{\alpha_0/2}(0,1) \text{ e } |W_{lb(j)}| \geq |W_{lb(i)}|, \forall i. \tag{10}$$

Retira-se a observação suspeita, e o teste deve ser repetido até que todas as observações suspeitas sejam identificadas.

**3 CONFIABILIDADE INTERNA E EXTERNA**

**3.1 Confiabilidade interna**

Ao se realizar um teste de hipóteses deve-se ter em mente que dois tipos de erros estão presentes. O erro tipo 1, representado por  $\alpha$ , é denominado de nível de significância de um teste de hipótese. Ele é definido como a probabilidade de rejeitar  $H_0$  (hipótese nula) quando na verdade  $H_0$  é verdadeira. O erro tipo 2, representado por  $\beta$ , se refere à probabilidade de aceitar  $H_0$  quando na realidade  $H_0$  é falsa, ou seja,  $H_a$  (hipótese alternativa) é verdadeira.

Normalmente, o valor de  $\alpha$  é definido e fixado, ficando como preocupação entender a probabilidade do erro tipo 2 de um dado teste de hipóteses. Em outras

palavras, fica a preocupação de construir um teste com proteção razoável de  $\beta$ .

O poder do teste, representado por  $\gamma$ , é a probabilidade de aceitar  $H_0$  quando realmente  $H_0$  é verdadeira. Ou seja, é rejeitar  $H_a$  quando de fato  $H_a$  é falsa. Em outras palavras, o poder do teste é o contrário do erro tipo 2:

$$\gamma = 1 - \beta \tag{11}$$

Dessa forma, é importante calcular o tamanho de  $\gamma$  para verificar qual é a proteção do teste de hipóteses com relação ao erro tipo 2 ( $\beta$ ).

O poder do teste  $\gamma$  depende de três fatores (TEUNISSEN, 2000):

- a) do tamanho de  $\alpha$ ;
- b) do número de graus de liberdade ( $q$ ) do teste (no caso do data snooping  $q = 1$ ). Não confundir  $q$  com o número de graus de liberdade ( $n - u$ ) de um ajustamento. O termo  $q$ , neste caso, é o número de observação que estão sendo testadas no teste de hipóteses.
- c) do parâmetro de não centralidade  $\lambda$  da distribuição de probabilidade de  $H_a$ .

Na tabela 1 são mostrados alguns valores típicos de  $\gamma$ , dados em função de  $\alpha$ ,  $q$  e  $\lambda$  (TEUNISSEN, 2000).

Tabela 1 – Valores do poder de  $\gamma$  para diferentes valores de  $\alpha$ ,  $q$  e  $\lambda$

$\alpha = 0,01$	$q = 1$	$q = 7$
$\lambda = 2$	0,1227	0,0415
$\lambda = 8$	0,5997	0,2710
$\lambda = 18$	0,9522	0,7430
$\alpha = 0,05$	$q = 1$	$q = 7$
$\lambda = 2$	0,2930	0,1378
$\lambda = 8$	0,8074	0,5017
$\lambda = 18$	0,9888	0,8946
$\alpha = 0,1$	$q = 1$	$q = 7$
$\lambda = 2$	0,4099	0,2272
$\lambda = 8$	0,8817	0,6287
$\lambda = 18$	0,9953	0,9413

Da tabela 1, pode-se observar que o poder do teste  $\gamma$ :

- aumenta junto com  $\alpha$ , para  $q$  e  $\lambda$  fixos;

- diminui com o aumento de  $q$ , para  $\alpha$  e  $\lambda$  fixos e
- aumenta junto com  $\lambda$ , para  $\alpha$  e  $q$  fixos.

Então, como  $\gamma$  depende de  $\alpha$ ,  $q$  e  $\lambda$ , existem três possibilidades para se construir um teste com proteção razoável do erro tipo 2. Pode-se, por exemplo, aumentar  $\alpha$ . Porém, o aumento de  $\alpha$  implica no aumento da probabilidade do erro tipo 1, o que não seria uma boa estratégia. Outro detalhe é que, usualmente, o tamanho de  $\alpha$  é escolhido como um valor fixo. Então, outra possibilidade é o decréscimo de  $q$ . Mas, também, o valor de  $q$  já é estipulado para cada caso, sendo uma característica (propriedade) do teste (no data snooping,  $q = 1$ ). Finalmente, deve-se então, aumentar o parâmetro de não centralidade  $\lambda$ . Mas como fazer isto? Tem-se, então, que estudar o comportamento de  $\lambda$ .

O valor de  $\lambda$  é dado por (TEUNISSEN, 2000):

$$\lambda = (C_{lb} \nabla_{lb})^T \left[ \sum_{Lb}^{-1} - \sum_{Lb}^{-1} A(A^T P A)^{-1} A^T \sum_{Lb}^{-1} \right] (C_{lb} \nabla_{lb}) \tag{12}$$

A equação (12) mostra que  $\lambda$  depende de:

- 1)  $\sum_{Lb}$  (a precisão das observações);
- 2) matriz  $A$  (geometria/configuração da rede) e
- 3)  $C_{lb} \nabla_{lb}$  (modelo de erro – representa o afastamento entre  $H_0$  e  $H_a$ ).

Da equação (12) pode-se, então, concluir que aumentando a precisão das observações aumenta-se  $\lambda$  e, conseqüentemente, o poder do teste  $\gamma$ . Isto implica na escolha de métodos e equipamentos adequados para o processo de coleta de dados. Outra influência é a geometria da rede, devendo-se planejar a mais adequada, bem como a redundância das observações. Quanto ao valor de  $C_{lb} \nabla_{lb}$ , ele é incógnito, impossibilitando o cálculo de  $\gamma$  na prática. Para o caso do data snooping,  $\nabla_{lb}$  é o valor do outlier na  $i$ -ésima observação  $lb_i$  e a matriz  $C_{lb}$  torna-se o vetor  $c_{lb(i)}$  (ver equação 5).

Pode-se, no entanto, assumir um valor representativo para  $C_{lb} \nabla_{lb}$  e calcular qual seria o valor de  $\lambda$  e, enfim, o valor do poder do teste  $\gamma$ . Por exemplo, no caso do data snooping, poderia se estipular um erro mínimo ( $\nabla_{lb(i)}$ ) que se desejaria que o teste identificasse, e calcularia o poder do teste  $\gamma$ . Se a probabilidade  $\gamma$  for menor que a esperada, é possível melhorar por meio de mudanças em  $\sum_{Lb}$  (precisão das observações) e na matriz  $A$  (geometria). Isto mostra a importância do estudo da confiabilidade no processo de planejamento de um levantamento.

Porém, normalmente nas aplicações geodésicas, se está mais interessado na dimensão do menor erro  $\nabla_{lb}$  que pode ser detectado com uma probabilidade estipulada  $\gamma$ ,

do que no poder do teste propriamente dito. Na prática assume-se um valor de referência para o poder do teste,  $\gamma = \gamma_0$  ( $\gamma_0 = 50\%$ ,  $70\%$ , usualmente,  $80\%$ ), e junto com  $\alpha$  e  $q$  obtém-se  $\lambda = \lambda_0$ , que simbolicamente é função de:

$$\lambda_0 = \lambda(\alpha, q, \gamma = \gamma_0). \quad (13)$$

Este valor pode ser encontrado em tabela (como na tabela 1 – mais detalhes em Teunissen (2000)). Usualmente, se adota o valor de  $\lambda_0=17,075$ , que está relacionado a  $\alpha_0 = 0,001$  e  $\gamma = 80\%$ . A partir de  $\lambda = \lambda_0$  pode-se calcular o modelo de erro  $C_{lb} \nabla_{lb}$ . Isto é obtido a partir da solução quadrática (TEUNISSEN, 2000):

$$\lambda_0 = \nabla_{lb}^T C_{lb}^T \sum_{Lb}^{-1} \sum_v \sum_{Lb}^{-1} C_{lb} \nabla_{lb} \quad (14)$$

para  $\nabla_{lb}$ . Uma vez que  $\nabla_{lb}$  é conhecido, o modelo de erro é calculado por:

$$\nabla_{Lb} = C_{lb} \nabla_{lb}. \quad (15)$$

O vetor  $\nabla_{Lb}$  é dito descrever a confiabilidade interna de  $H_0$  com respeito a  $H_a$ . Em outras palavras, pode-se afirmar que a “*confiabilidade interna é uma dimensão do modelo de erro que pode ser detectada com uma probabilidade  $\gamma = \gamma_0$* ”. Mas como calcular o tamanho de  $\nabla_{lb}$  pela equação (14)? Infelizmente, a equação (14) não possui solução única para  $\nabla_{lb}$ . Considerando o caso do data snooping, a matriz  $C_{lb}$  reduz-se a um vetor  $c_{lb(i)}$  de dimensão  $(n \times 1)$ , e o vetor  $\nabla_{lb}$  de dimensão  $(q \times 1)$  reduz-se a um escalar  $\nabla_{lb(i)}$  (ver equação 5). Então, para o caso do data snooping a equação (14) pode ser escrita como (TEUNISSEN, 2000):

$$\lambda_0 = c_{lb(i)}^T \sum_{Lb}^{-1} \sum_v \sum_{Lb}^{-1} c_{lb(i)} (\nabla_{lb(i)})^2 \quad (16)$$

Logo, a solução em termos de  $\nabla_{lb(i)}$  é dada por (TEUNISSEN, 2000):

$$|\nabla_{lb(i)}| = \sqrt{\frac{\lambda_0}{c_{lb(i)}^T \sum_{Lb}^{-1} \sum_v \sum_{Lb}^{-1} c_{lb(i)}}}. \quad (17)$$

O escalar  $|\nabla_{lb(i)}|$  é o menor erro (outlier) que pode ser detectado na  $i$ -ésima observação usando o teste data snooping, com probabilidade  $\gamma = \gamma_0$ . Desta forma, a equação (17) mostra que o **Menor Erro Detectável (MDB – Minimal Detectable Bias)** depende, primeiramente, da escolha de  $\gamma = \gamma_0$ ,  $\alpha$  e  $q$ . Para o caso do data snooping o nível de significância é  $\alpha_0$  e  $q = 1$ . O valor de  $\alpha_0$  do data snooping deve estar relacionado adequadamente com o  $\alpha$  do Teste Global ( $q = n - u$ ). Esta

relação é mostrada pelo parâmetro de não centralidade  $\lambda_0$  que deve ser o mesmo valor tanto para o data snooping quanto para o Teste Global. Desta forma, tem-se a seguinte relação:

$$\lambda_0 = \lambda(\alpha, q = n-u, \gamma = \gamma_0) = \lambda(\alpha_0, q = 1, \gamma = \gamma_0) \quad (18)$$

Teste Global                      Data Snooping

Ou seja, adotado um valor para  $\alpha_0$  obtém-se o valor de  $\lambda_0$ . Em seguida, obtém-se o valor de  $\alpha$ , tal que o valor de  $\lambda_0$  seja mantido o mesmo. Ou vice-versa iniciando com o valor de  $\alpha$ . Uma relação algébrica aproximada entre  $\alpha$  e  $\alpha_0$  é dada por (CHEN et al., 1987):

$$\alpha \cong 2 n \alpha_0. \quad (19)$$

Outra dependência da equação (20) de  $\nabla_{lb(i)}$  é quanto à precisão das observações (contidas em  $\sum_{Lb}^{-1}$  e  $\sum_v$ ) e a geometria/configuração da rede (matriz  $A$  – contida em  $\sum_v$ ).

Em muitas aplicações práticas, a matriz variância-covariância das observações é diagonal, ou seja, ignora-se ou não existe correlação entre as observações. Neste caso, a equação (17) pode ser simplificada para:

$$|\nabla_{lb(i)}| = \sigma_{lb(i)} \sqrt{\frac{\lambda_0}{[1 - (\sigma_{la(i)}^2 / \sigma_{lb(i)}^2)]}} \quad (20)$$

onde  $\sigma_{lb(i)}^2$  é a precisão da  $i$ -ésima observação e  $\sigma_{la(i)}^2$  é a precisão da  $i$ -ésima observação ajustada (ver Gemael, 1994).

A equação (20) mostra que o menor erro detectável ( $|\nabla_{lb(i)}|$ ) na  $i$ -ésima observação é maior se o valor de  $\sigma_{la(i)}^2$  for próximo de  $\sigma_{lb(i)}^2$ , e que  $|\nabla_{lb(i)}|$  é menor quanto menor for o valor de  $\sigma_{la(i)}^2$  em comparação a  $\sigma_{lb(i)}^2$ . Quanto menor o valor de  $|\nabla_{lb(i)}|$  melhor é a confiabilidade interna do modelo, ou seja, da rede.

### 3.1.1 Confiabilidade interna na etapa de planejamento

A equação (20) só pode ser calculada após o ajustamento ter sido realizado.

Em muitos casos, antes mesmo do levantamento de campo ser realizado, se tem o conhecimento *a priori* de como será a geometria da rede de pontos a ser levantada e também se tem uma noção de como será a precisão das observações, por meio do conhecimento dos

equipamentos que poderão ser usados, dos métodos a ser aplicados, entre outros fatores.

Neste sentido, seria importante no processo de planejamento que se tivesse condições de se calcular um valor esperado para  $|\nabla_{lb(i)}|$  antes mesmo que o levantamento fosse executado. Isto possibilitaria analisar a confiabilidade interna da rede, ou pelo menos se ter uma noção, antes mesmo que as observações fossem coletadas. Mas este objetivo não pode ser alcançado da forma em que está definida a equação (20) (pois,  $\sigma_{la(i)}^2$  é desconhecida antes do ajustamento). Porém, em Teunissen (2000) é mostrado um desenvolvimento da equação (20) de tal forma que se pode obter uma confiabilidade interna média para a  $i$ -ésima observação, antes do levantamento de campo ser realizado, ou seja, na fase de planejamento.

O termo adimensional da equação (20):

$$r_{(i)} = 1 - (\sigma_{la(i)}^2 / \sigma_{lb(i)}^2), \quad (21)$$

denominado agora de  $r_{(i)}$  é conhecido como sendo o  $i$ -ésimo número de redundância local. Note que, como  $0 \leq \sigma_{la(i)}^2 \leq \sigma_{lb(i)}^2$ , logo, o número de redundância local  $r_{(i)}$  sempre está compreendido no seguinte intervalo:

$$0 \leq r_{(i)} \leq 1, \quad (22)$$

A razão pela qual  $r_{(i)}$  é chamado de  $i$ -ésimo número de redundância local é devido a seguinte propriedade:

$$\sum_{i=1}^n r_{(i)} = \sum_{i=1}^n 1 - (\sigma_{la(i)}^2 / \sigma_{lb(i)}^2) = n - u \quad (23)$$

onde,  $n$  é o número de observações e  $u$  é o número de parâmetros envolvidos no ajustamento. Ou seja, a soma de todos os números de redundância local é igual ao número de observações redundantes no modelo, em outras palavras, igual ao grau de liberdade ( $n - u$ ) do ajustamento. Da equação (21) pode-se, então, obter um valor médio  $\bar{r}$  da seguinte forma:

$$\bar{r} = \frac{\sum_{i=1}^n r_{(i)}}{n} = \frac{n - u}{n} \quad (24)$$

Logo, substituindo na equação (20) o  $i$ -ésimo número de redundância local por  $\bar{r}$  (redundância média), pode-se obter uma aproximação para o MDB da  $i$ -ésima observação ( $|\nabla_{lb(i)}|$ ):

$$|\nabla_{lb(i)}| \cong \sigma_{lb(i)} \sqrt{\frac{\lambda_o}{(n-u)/n}} \quad (25)$$

A equação (25) mostra que, conhecendo-se *a priori* ou pelo menos tendo uma noção da precisão da  $i$ -ésima observação, bem como do número de observações e de parâmetros (ou seja, a geometria da rede), pode-se calcular um valor aproximado para o menor erro que pode ser detectado no data snooping com uma probabilidade  $\gamma = \gamma_o$ , isto tudo antes mesmo que o levantamento de campo fosse realizado. Isto pode ser mais um parâmetro a ser analisado na etapa de planejamento. Se os valores de  $|\nabla_{lb(i)}|$  não forem os que se esperava, podem-se adotar as medidas necessárias, que, por exemplo, vão desde a escolha de outros equipamentos até a mudança da geometria da rede planejada. É claro, que outros fatores podem ser modificados, variando para cada caso.

#### 4 CONFIABILIDADE EXTERNA

A confiabilidade interna foi apresentada e definida de forma geral como sendo o modelo de erro que pode ser detectado por um teste de hipótese com uma probabilidade  $\gamma = \gamma_o$ . Determinado  $\nabla_{lb}$  a confiabilidade interna é descrita de maneira geral pelo vetor ( $n \times 1$ ):

$$\nabla_{Lb} = C_{lb} \nabla_{lb} \quad (26)$$

No caso do data snooping, determinado o menor erro detectável ( $\nabla_{lb(i)}$ ) na  $i$ -ésima observação, a equação (26) reduz-se para:

$$\nabla_{Lb(i)} = c_{lb(i)} |\nabla_{lb(i)}| \quad (27)$$

onde:  $c_{lb(i)}^T = [0 \dots 0 \dots 1 \dots 0 \dots 0]$

Logo, o vetor  $\nabla_{Lb(i)}$  para o caso do data snooping será dado por:

$$\nabla_{Lb(i)}^T = \left[ 0 \dots 0 \dots |\nabla_{lb(i)}| \dots 0 \dots 0 \right] \quad (28)$$

Ou seja, será um vetor em que apenas na posição da  $i$ -ésima observação que está sendo testada não possui o valor nulo, e sim possui o menor erro detectável  $|\nabla_{lb(i)}|$ .

Logo, definido a confiabilidade interna a próxima etapa é verificar a confiabilidade externa. Esta é definida como sendo a "influência do modelo de erro ( $\nabla_{Lb} = C_{lb} \nabla_{lb}$ ) sobre os resultados finais de um ajustamento". Por exemplo, para o caso do data snooping, a confiabilidade externa mostra a influência do MDB ( $|\nabla_{lb(i)}|$ ) da  $i$ -ésima observação nos parâmetros estimados.

Ou seja, é importante saber a influência nos parâmetros de um possível erro não detectado. E o erro mínimo que pode ser detectado, advindo da confiabilidade interna, fornece a possibilidade de analisar esta influência.

Enfim, a confiabilidade externa é obtida utilizando o mesmo estimador dos mínimos quadrados, apenas substituindo o vetor  $L_b$  pelo MDB ( $\nabla_{L_b}$ ). Logo, a confiabilidade externa é dada por (TEUNISSEN, 2000):

$$\nabla X_{(i)} = (A^T P A)^{-1} [A^T P \nabla_{L_b}] \quad (29)$$

Então,  $\nabla X_{(i)}$  é o vetor de dimensão (ux1) que descreve o erro esperado nos parâmetros causados pelo possível modelo de erro ( $\nabla_{L_b}$ ) não detectado no teste estatístico.

Para o caso do data snooping, a equação 29, ou seja, a influência de  $|\nabla_{l_{b(i)}}|$  no vetor dos parâmetros X pode ser escrita da seguinte forma (TEUNISSEN, 2000):

$$\nabla X_{(i)} = (A^T P A)^{-1} [A^T P (c_{l_{b(i)}} |\nabla_{l_{b(i)}}|)] \quad (30)$$

Neste caso,  $\nabla X_{(i)}$  é o vetor de dimensão (ux1) que descreve a influência do menor erro detectável na i-ésima observação no vetor dos parâmetros X. Uma abordagem mais ampla pode ser encontrada em Teunissen (2000).

### 5 ANÁLISE DA CONFIABILIDADE EM UMA REDE DE NIVELAMENTO NA ETAPA DE PLANEJAMENTO

A rede de nivelamento utilizada é apresentada em Gemael (1994, p.147). Ela é composta por 16 linhas de nivelamento, com 11 altitudes a serem ajustadas e 1 conhecida (Figura 1)

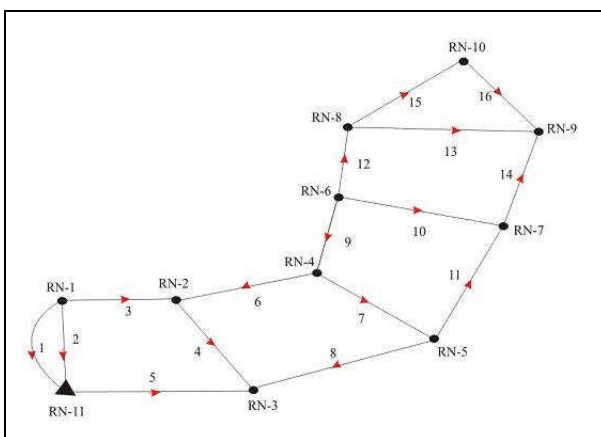


Figura 1 – Rede de nivelamento utilizada no estudo (adaptada de Gemael, 1994).

A tabela 2 apresenta os valores do comprimento das linhas niveladas (K) associadas a cada observação de desnível. Este comprimento será útil para adotar critério para ponderação das observações.

Tabela 2 – Comprimento de cada linha nivelada

Observação ( $l_{b(i)}$ )	Comprimento (K) (em km)
1	72
2	40
3	44
4	20
5	61
6	26
7	59
8	60
9	83
10	26
11	74
12	248
13	111
14	67
15	108
16	147

Obviamente, esta rede de nivelamento já foi levantada e possui o valor dos desníveis observados, e, naturalmente, a confiabilidade interna e externa após o ajustamento da rede podem ser calculadas. Porém o foco deste artigo é abordar a confiabilidade na etapa de planejamento do levantamento, necessitando somente da geometria da rede e da possível precisão das observações. No caso da rede de nivelamento é bastante usual ponderar as observações de desníveis com relação ao comprimento da linha nivelada, sendo esta estratégia adotada neste estudo.

O cálculo da confiabilidade interna e externa desta rede após ter sido realizado o Ajustamento da mesma, já foi realizado, apresentado e analisado por Monico et al. (2006).

#### 4.1 Análise da confiabilidade interna e externa da rede de nivelamento na etapa de planejamento

Para o cálculo da confiabilidade interna antes da execução do levantamento é utilizada a equação (25). Para a aplicação desta equação é necessário adotar a possível precisão das observações e o número de observações e de parâmetros da rede. Além disso, também se deve adotar um valor para o poder do teste  $\gamma$  e para o nível de significância  $\alpha_0$ , para que se obtenha um valor para o parâmetro de não centralidade  $\lambda_0$ . Foi adotado o valor de 0,80 e 0,001 para o poder do teste e para o nível de significância, respectivamente, valores normalmente adotados na literatura. Com estes valores, o valor de  $\lambda_0$  é de 17,075 (verificado em tabela).

Primeiramente, para realizar o estudo, adotou-se o valor da precisão de cada observação como sendo de  $\sigma_{l_{b(i)}} = 6 \text{ mm } \sqrt{K}$ , onde K (em km) é o comprimento de cada linha nivelada (tabela 2). Aplicando a equação (25) calculou-se a confiabilidade interna. A tabela 3 mostra os

valores calculados para a confiabilidade interna na fase de planejamento, ou seja, os valores do menor erro ( $|\nabla_{lb(i)}|$ ) que poderia ser detectado em cada observação pelo data snooping, considerando  $\gamma = \gamma_o = 80\%$  e  $\alpha_o = 0,001$ .

Tabela 3 – Confiabilidade interna na fase de planejamento (para  $\sigma_{lb(i)} = 6 \text{ mm } \sqrt{K}$ )

CONFIABILIDADE INTERNA – Fase de Planejamento		
Considerando: $n = 16; u = 10; \sigma_{lb(i)} = 6 \text{ mm } \sqrt{K}$ ;		
$\gamma_o = 80\% \alpha_o = 0,001$		
Observação ( $lb(i)$ )	Precisão ( $\sigma_{lb(i)}$ ) (m)	$ \nabla_{lb(i)} $ (m)
1	0,050	<b>0,343</b>
2	0,037	<b>0,256</b>
3	0,039	<b>0,268</b>
4	0,026	<b>0,181</b>
5	0,046	<b>0,316</b>
6	0,030	<b>0,206</b>
7	0,046	<b>0,310</b>
8	0,046	<b>0,313</b>
9	0,054	<b>0,368</b>
10	0,030	<b>0,206</b>
11	0,051	<b>0,348</b>
12	0,094	<b>0,637</b>
13	0,063	<b>0,426</b>
14	0,049	<b>0,331</b>
15	0,062	<b>0,420</b>
16	0,072	<b>0,490</b>

Pela tabela 3 verifica-se que os valores do menor erro detectável (MDB -  $|\nabla_{lb(i)}|$ ) calculados na fase de planejamento, considerando  $\sigma_{lb(i)} = 6 \text{ mm } \sqrt{K}$ , ficaram entre 0,18 e 0,63 m. Se estivesse em uma situação real de planejamento, se o projetista considerar esses valores maiores do que o esperado, ele deveria estudar uma forma de melhorar a confiabilidade interna da rede. Pode-se melhorar a precisão das observações e/ou estudar outra geometria para rede, aumentar o número de observações, etc. Note que os valores de confiabilidade interna para cada observação da tabela 3 são melhores quanto melhor é a precisão.

Por exemplo, se o projetista resolvesse melhorar a precisão das observações (adotando outro equipamento, por exemplo) e passar a possuir uma precisão de  $\sigma_{lb(i)} = 2 \text{ mm } \sqrt{K}$  ao invés de  $\sigma_{lb(i)} = 6 \text{ mm } \sqrt{K}$ , poderá se obter melhores valores de confiabilidade interna da rede. Isto pode ser verificado na tabela 4 que apresenta os mesmos parâmetros da tabela 3, só que agora considerando  $\sigma_{lb(i)} = 2 \text{ mm } \sqrt{K}$ .

Tabela 4 – Confiabilidade interna na fase de planejamento (para  $\sigma_{lb(i)} = 2 \text{ mm } \sqrt{K}$ )

CONFIABILIDADE INTERNA – Fase de Planejamento		
Considerando: $n = 16; u = 10; \sigma_{lb(i)} = 2 \text{ mm } \sqrt{K}$ ;		
$\gamma_o = 80\% \alpha_o = 0,001$		
Observação ( $lb(i)$ )	Precisão ( $\sigma_{lb(i)}$ ) (m)	$ \nabla_{lb(i)} $ (m)
1	0,0169	<b>0,114</b>
2	0,0126	<b>0,085</b>
3	0,0132	<b>0,089</b>
4	0,0089	<b>0,068</b>
5	0,0156	<b>0,105</b>
6	0,0101	<b>0,068</b>
7	0,0153	<b>0,103</b>
8	0,0154	<b>0,104</b>
9	0,0182	<b>0,122</b>
10	0,0101	<b>0,068</b>
11	0,0172	<b>0,116</b>
12	0,0314	<b>0,212</b>
13	0,0210	<b>0,142</b>
14	0,0163	<b>0,110</b>
15	0,0207	<b>0,140</b>
16	0,0242	<b>0,490</b>

Note que os valores do MDB da tabela 4 são menores do que os da tabela 3. O valor mínimo foi de 0,068 m e o máximo de 0,212. Ou seja, melhorando a precisão das observações melhora-se a confiabilidade interna da rede. Pode-se, pensar então, na fase de planejamento, em estudar o comportamento da confiabilidade interna para diferentes valores de precisão. Neste sentido, a figura 5 mostra os valores do menor erro detectável para cada observação, considerando como precisão das observações os casos de:  $\sigma_{lb(i)} = 1 \text{ mm } \sqrt{K}$ ,  $\sigma_{lb(i)} = 2 \text{ mm } \sqrt{K}$ ,  $\sigma_{lb(i)} = 3 \text{ mm } \sqrt{K}$ ,  $\sigma_{lb(i)} = 4 \text{ mm } \sqrt{K}$ ,  $\sigma_{lb(i)} = 5 \text{ mm } \sqrt{K}$  e  $\sigma_{lb(i)} = 6 \text{ mm } \sqrt{K}$ .

Logo, a figura 2 apresenta os valores do MBD ( $|\nabla_{lb(i)}|$ ) para cada observação considerando os casos de precisão citados anteriormente.



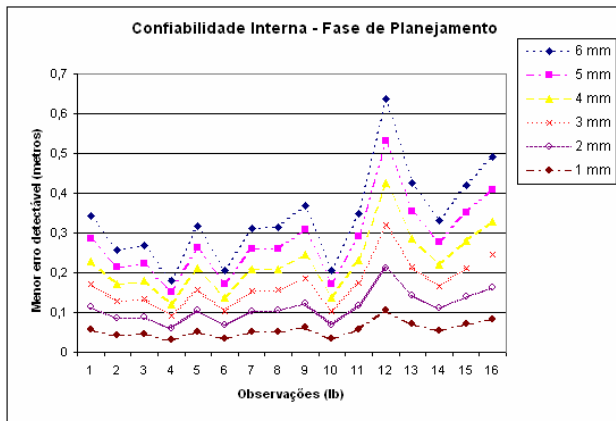


Figura 2 – MDB ( $|\nabla_{lb(i)}|$ ) para diferentes valores de precisão – fase de planejamento

Novamente, como se espera, verifica-se na figura 2 que quanto melhor a precisão das observações melhor será a confiabilidade interna da rede. Por exemplo, enquanto o MDB da observação 13 é de 0,63 m para o caso de precisão de 6 mm  $\sqrt{K}$ , esse mesmo valor é reduzido para 0,10 m para o caso de 1 mm  $\sqrt{K}$ .

Após o cálculo da confiabilidade interna pode-se prosseguir com o estudo da confiabilidade externa da rede, ainda na fase de planejamento do levantamento. No exemplo que está sendo adotado, onde se está considerando a confiabilidade interna para o caso do data snooping, para cada MDB tem-se uma confiabilidade externa. Em outras palavras, a confiabilidade externa no caso adotado mostra a influência de um erro mínimo detectável na i-ésima observação no vetor dos parâmetros, que neste exemplo são as altitudes dos pontos. Dessa forma, como se tem 16 observações pode-se calcular 16 vetores de confiabilidade externa, ou seja, a influência de de cada MDB de cada observação nos parâmetros (altitudes).

Para exemplificar, considerem-se os dados de confiabilidade interna da tabela 3 para  $\sigma_{lb(i)} = 6 \text{ mm } \sqrt{K}$ . Será calculado a influência no vetor dos parâmetros do menor valor de  $|\nabla_{lb(i)}|$  que ocorreu na observação 4, e posteriormente, do maior valor que ocorreu na observação 12. Então, tem-se:

- Para  $|\nabla_{lb(4)}| = 0,18 \text{ m (o menor)}$ :

$$\nabla_{Lb(4)}^T = \begin{bmatrix} 0 & 0 & \dots & |\nabla_{lb(4)}| & \dots & 0 & 0 \\ 1 & 2 & & 4 & & n-1 & n \end{bmatrix} =$$

$$\nabla_{Lb(4)}^T = \begin{bmatrix} 0 & 0 & \dots & 0,18 & \dots & 0 & 0 \\ 1 & 2 & & & & n-1 & n \end{bmatrix}$$

Logo, a confiabilidade externa neste caso será:

$$\nabla X = (A^T P A)^{-1} [A^T P \nabla_{Lb(4)}] =$$

$$\nabla X = \begin{bmatrix} \nabla H_1 \\ \nabla H_2 \\ \nabla H_3 \\ \nabla H_4 \\ \nabla H_5 \\ \nabla H_6 \\ \nabla H_7 \\ \nabla H_8 \\ \nabla H_9 \\ \nabla H_{10} \end{bmatrix} = \begin{bmatrix} 0,027 \\ 0,073 \\ 0,064 \\ 0,046 \\ 0,0009 \\ 0,024 \\ 0,018 \\ 0,020 \\ 0,019 \\ 0,020 \end{bmatrix} \text{ m}$$

- Para  $|\nabla_{lb(12)}| = 0,63 \text{ m (o maior)}$ :

$$\nabla_{Lb(12)}^T = \begin{bmatrix} 0 & 0 & \dots & |\nabla_{lb(12)}| & \dots & 0 & 0 \\ 1 & 2 & & 12 & & n-1 & n \end{bmatrix} =$$

$$\nabla_{Lb(12)}^T = \begin{bmatrix} 0 & 0 & \dots & 0,63 & \dots & 0 & 0 \\ 1 & 2 & & & & n-1 & n \end{bmatrix}$$

Logo, a confiabilidade externa será:

$$\nabla X = (A^T P A)^{-1} [A^T P \nabla_{Lb(12)}] =$$

$$\nabla X = \begin{bmatrix} \nabla H_1 \\ \nabla H_2 \\ \nabla H_3 \\ \nabla H_4 \\ \nabla H_5 \\ \nabla H_6 \\ \nabla H_7 \\ \nabla H_8 \\ \nabla H_9 \\ \nabla H_{10} \end{bmatrix} = \begin{bmatrix} 0,0002 \\ 0,0006 \\ 0,0005 \\ 0,002 \\ 0,004 \\ 0,017 \\ 0,017 \\ 0,239 \\ 0,120 \\ 0,189 \end{bmatrix} \text{ m}$$

Os dois vetores  $\nabla X$  mostram o erro que se pode cometer nas altitudes após o ajustamento caso um erro não detectado  $|\nabla_{lb(4)}| = 0,18 \text{ m}$  e  $|\nabla_{lb(12)}| = 0,63 \text{ m}$  fosse cometido, respectivamente. Eles descrevem a confiabilidade externa da rede calculada na fase de planejamento. Quanto melhor for a confiabilidade interna melhor será a externa. Então, adotando estratégias para melhorar a confiabilidade interna estará melhorando, também, a confiabilidade externa.

## 6 CONSIDERAÇÕES FINAIS

O principal foco deste artigo foi apresentar dentro do contexto do Ajustamento de Observações o cálculo da confiabilidade na fase de planejamento do levantamento de campo. Neste sentido, a teoria sobre confiabilidade (interna e externa) e um simples exemplo de aplicação para o caso de uma rede de nivelamento foram apresentados. Pretendeu-se mostrar que a confiabilidade também é uma etapa que pode ser considerada dentro da etapa de planejamento, como auxílio adicional na definição das estratégias s serem empregadas. Como visto, é importante a inclusão da confiabilidade, pois, ela está ligada com os testes estatísticos a serem aplicados no Ajustamento. Explicitando de maneira simples, a confiabilidade mostra a potencialidade dos testes estatísticos.

Nos experimentos apenas simulações com relação a diferentes precisões foram analisadas. É importante salientar que simulações de diferentes situações em termos de geometria da rede também poderiam ter sido realizadas, sendo um possível tema de trabalho futuro.

Este artigo é uma continuação dos trabalhos já iniciados em Monico et al. (2006), e novos trabalhos deverão ser explorados neste tema.

## REFERÊNCIAS

BAARDA, W.. A testing procedure for use in geodetic networks. **Netherlands Geodetic Commission – Publication on Geodesy – News Series**, Delft, v.2, n.5, 1968.

CHEN, Y.Q.; KAVOURAS, M.; CHRZANOWSKI, A. A strategy for detection of outlying observations in measurements of high precision. **The Canadian Surveyor**, Ottawa, v.41, n.4, p.529-540, winter 1987.

GEMAEL, C.. **Introdução ao ajustamento de observações: aplicações geodésicas**. Curitiba: Ed. UFPR, 1994. 319p.

MONICO, J.F.G.; MATSUOKA, M.T.; SAPUCCI, L.F.. Confiabilidade interna e externa em aplicações geodésicas: Exemplo de uma rede de nivelamento, **Geodésia online**, n.2, 2006.

TEUNISSEN, P.J.G.. **Testing Theory: an introduction**. 2000. Editora: Delft University Press. Delft, Holanda.

TEUNISSEN, P.J.G.. **Quality Control and GPS**. In: Teunissen, P.J.G.; Kleusberg, A. *GPS for Geodesy*. 2.ed.: Springer Verlag, 1998. p. 271-318.