

Pesquisas em Geociências

<http://seer.ufrgs.br/PesquisasemGeociencias>

Uma Técnica Multivariada de Detecção e Identificação de Outliers em Fototriangulação

Quintino Dalmolin, Jair Mendes Marques

Pesquisas em Geociências, 21 (2): 100-106, set./dez., 1994.

Versão online disponível em:

<http://seer.ufrgs.br/PesquisasemGeociencias/article/view/21242>

Publicado por

Instituto de Geociências



Portal de Periódicos UFRGS

UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL

Informações Adicionais

Email: pesquisas@ufrgs.br

Políticas: <http://seer.ufrgs.br/PesquisasemGeociencias/about/editorialPolicies#openAccessPolicy>

Submissão: <http://seer.ufrgs.br/PesquisasemGeociencias/about/submissions#onlineSubmissions>

Diretrizes: <http://seer.ufrgs.br/PesquisasemGeociencias/about/submissions#authorGuidelines>

Data de publicação - set./dez., 1994.

Instituto de Geociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

Uma Técnica Multivariada de Detecção e Identificação de *Outliers* em Fototriangulação

QUINTINO DALMOLIN¹ E JAIR MENDES MARQUES²

¹ Departamento de Geociências; UFPR; Caixa Postal 19011; CEP 81.531-990, Curitiba, PR, Brasil

² Departamento de Estatística; UFPR; CEP 81.531-990, Curitiba, PR, Brasil

(Recebido em 20/05/94. Aceito para publicação em 06/12/94.)

Abstract - This paper presents a technique of multivariate statistical analysis for detecting and identifying multiple outliers in phototriangulation. A simulation involving a block of six photographs is presented as well as the obtained results by the multivariate methods, during this simulation outliers are imposed in three observations.

Resumo - Este artigo apresenta uma técnica de análise estatística multivariada na detecção e identificação de múltiplos *outliers* em fototriangulação. Uma simulação envolvendo um bloco de seis fotografias é apresentada bem como os resultados obtidos pelo método multivariado quando são impostos erros grosseiros em três observações.

INTRODUÇÃO

Os estudos de detecção e identificação de *outliers* em fototriangulação, têm sido em geral realizados com base nos resíduos padronizados. O método mais utilizado é conhecido como *data snooping* de Baarda (Berberan, 1992; El-Hakim, 1981; Firkowski, 1986; Kubik & Wang, 1991; Mitishita, 1986; Ou, 1989).

Nos métodos que utilizam a análise dos resíduos padronizados, a suposição fundamental é que as observações contaminadas apresentam resíduos padronizados de magnitudes significativamente altas. Porém, essa suposição nem sempre é verdadeira; constituindo-se então como uma dificuldade desse método como forma de detectar e identificar *outliers*. Esse problema surge em decorrência do ajustamento pelo método dos mínimos quadrados, já que no processo de ajustamento o erro de uma observação irá contaminar outras observações, no fenômeno conhecido em análise estatística como efeito *smearing*. Como consequência, fica comprometida a sensibilidade dos resíduos para análise dos erros grosseiros nas observações. A detecção torna-se ainda mais complicada e menos confiável na presença de múltiplos *outliers*. Um outro método envolvendo resíduos padronizados, que tem sido utilizado é o método de Pope (Pope, 1976) que consiste basicamente numa simplificação do método de Baarda.

Nas análises envolvendo múltiplos *outliers*, o método mais utilizado envolve a estimação robusta, sendo muito utilizado, o *danish method* (Berberan, 1992; Clerici & Harris, 1980; Crosilla, 1983; Cross & Price, 1985; Davies & Gather, 1993; Patterson, 1985; Silva, 1987).

Com o crescente aumento da velocidade de processamento dos computadores, as técnicas multivariadas surgem como alternativa interessante no estudo de detecção e identificação de *outliers*. Particularmente o que se pretende com esse trabalho é fornecer um método alternativo para a detecção e identificação de múltiplos *outliers* em fototriangulação. O método a ser proposto utiliza como ferramenta principal o método multivariado conhecido como

Análise de Componentes Principais (ACP) (Aduol, 1987; Dalmolin & Marques, 1993; Jackson & Mudholkar, 1979, Johnson & Wichern, 1988; Morrison, 1976). O uso da ACP permitirá inicialmente construir um teste de hipótese envolvendo a detecção de *outliers*, e em seguida pode-se ainda, através da ACP, separar as prováveis observações contaminadas por erros. De posse dessas observações suspeitas, pode-se construir um teste t que finalmente permitirá identificar os *outliers*.

Além do desenvolvimento teórico do método proposto, apresenta-se um experimento envolvendo a detecção e identificação de *outliers* em fototriangulação. Para isso simulou-se um bloco de 6 fotografias, com 34 pontos aerotriangulados e 150 pontos imagens, impondo-se que 3 das fotocoordenadas estejam contaminadas de erros grosseiros.

FUNDAMENTOS TEÓRICOS DO MÉTODO

Para mostrar o desenvolvimento do método, será utilizado para o ajustamento, o conhecido modelo Gauss-Markov

$$\underline{Y} = \underline{AX} + \underline{V} \quad (1)$$

onde: \underline{Y} = vetor das observações (nx1);
 \underline{A} = matriz do modelo (nxu), de rank u;
 \underline{X} = vetor dos parâmetros (ux1);
 \underline{V} = vetor dos resíduos (nx1).

Para o modelo de Gauss-Markov (1), tem-se que

$$E(\underline{Y}) = \underline{AX} \quad (2)$$

$$E(\underline{V}) = \underline{0} \quad (3)$$

$$\Sigma_V = \Sigma_V = \sigma^2 P^{-1} \quad (4)$$

onde: P = matriz dos pesos das observações (nxn);
 σ^2 = variância da unidade de peso das observações.

O modelo estimado de (1) pode ser

$$\underline{Y} = A\underline{\hat{X}} + \underline{\hat{V}} \tag{5}$$

onde: $\underline{\hat{X}}$ = estimativa de \underline{X} ;

$\underline{\hat{V}}$ = estimativa de \underline{V} .

Do modelo estimado (5), tem-se que

$$\underline{\hat{V}} = \underline{Y} - A\underline{\hat{X}} \tag{6}$$

onde a aplicação do método dos mínimos quadrados possibilita a obtenção do vetor dos parâmetros estimados

$$\underline{\hat{X}} = (A'PA)^{-1} A'P\underline{Y} \tag{7}$$

sendo $A' = A'$ (transposta da matriz A).

Seja o vetor dos resíduos estimados

$$\underline{\hat{V}} = [\hat{v}_1 \hat{v}_2 \dots \hat{v}_n]' \tag{8}$$

onde $\hat{v}_1, \hat{v}_2, \dots, \hat{v}_n$, são n variáveis correlacionadas com

$$E(\underline{\hat{V}}) = \underline{0} \text{ e } \underline{\hat{\Sigma}}_v = \sigma^2 [P^{-1} - A(A'PA)^{-1}A'] \tag{9}$$

O vetor das componentes principais (Aduol, 1987) de $\underline{\hat{V}}$ será então

$$\underline{\hat{c}} = E_v' \cdot \underline{\hat{V}} \tag{10}$$

onde E_v é a matriz nxn cujas colunas são os autovetores normalizados dos resíduos estimados. Assim, $\underline{\hat{c}}$ e E_v podem ser escritos como

$$\underline{\hat{c}} = \begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \vdots \\ \hat{c}_n \end{bmatrix} \text{ e } E_v = \begin{bmatrix} e_{11} & e_{12} & \dots & e_{n1} \\ e_{12} & e_{22} & \dots & e_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1n} & e_{2n} & \dots & e_{nn} \end{bmatrix} \tag{11}$$

Pode-se mostrar (Aduol, 1987) que

$$E(\underline{\hat{c}}) = \underline{0} \tag{12}$$

$$\underline{\hat{\Sigma}}_c = \Lambda_v = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \tag{13}$$

sendo $\lambda_1, \lambda_2, \dots, \lambda_n$, os n autovalores da matriz dos resíduos estimados.

A padronização dos resíduos estimados conduz ao vetor

$$\underline{\hat{z}} = (\underline{\hat{\Sigma}}_v^{-1/2}) \cdot \underline{\hat{V}} \tag{14}$$

sendo $\underline{\hat{\Sigma}}_v$ a matriz diagonal nxn cujos elementos diagonais são as variâncias obtidas da matriz, $\underline{\hat{\Sigma}}_v$ dos resíduos estimados.

Para o vetor $\underline{\hat{z}}$ tem-se que

$$E(\underline{\hat{z}}) = \underline{0} \tag{15}$$

e

$$\underline{\hat{\Sigma}}_z = \underline{\hat{\rho}}_v = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \dots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \dots & 1 \end{bmatrix} \tag{16}$$

ou seja, a matriz covariância de $\underline{\hat{z}}$ é a própria matriz correlação de $\underline{\hat{V}}$.

No ajustamento do modelo (1), sendo $k = n - u$ o número de graus de liberdade; tem-se então entre os n autovalores: $\lambda_1, \lambda_2, \dots, \lambda_n$, da matriz $\underline{\hat{\rho}}_v$, k autovalores não-nulos. A matriz dos k autovetores normalizados correspondentes a esses k autovalores será denotada por E_{zk} . Portanto, o vetor das componentes principais do vetor $\underline{\hat{z}}$ será determinado pela relação

$$\underline{\hat{y}} = E_{zk}' \cdot \underline{\hat{z}} \tag{17}$$

onde:

$$E(\underline{\hat{y}}) = \underline{0} \tag{18}$$

$$\underline{\hat{\Sigma}}_y = \Lambda_{zk} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_k \end{bmatrix} \tag{19}$$

O vetor das k componentes principais padronizadas do vetor $\underline{\hat{y}}$ será então

$$\underline{\hat{s}} = \Lambda_{zk}^{-1/2} \cdot \underline{\hat{y}} \tag{20}$$

que pode ser colocado na forma

$$\underline{\hat{s}} = \Lambda_{zk}^{-1/2} \cdot E_{zk}' \cdot \underline{\hat{\Sigma}}_v^{-1/2} \cdot \underline{\hat{V}} \tag{21}$$

onde:

$$E(\underline{\hat{s}}) = \underline{0} \tag{22}$$

e

$$\underline{\hat{\Sigma}}_s = I_k \tag{23}$$

sendo I_k uma matriz identidade $k \times k$.

Como o vetor \hat{s} representa as componentes principais padronizadas, pode-se determinar, sob a hipótese de normalidade, a estatística qui-quadrado

$$\chi^2 = \hat{s}_1^2 + \hat{s}_2^2 + \dots + \hat{s}_k^2 = \sum_{i=1}^k \hat{s}_i^2 \quad (24)$$

com $\nu = n - u - 1 = k - 1$ graus de liberdade

Voltando ao modelo estimado de Gauss-Markov (5), o vetor \hat{Y} pode ser colocado na forma

$$\hat{Y} = Q_v P Y \quad (25)$$

onde

$$Q_v = P^{-1} - A(A'PA)^{-1}A' \quad (26)$$

Substituindo (25) em (21) resulta

$$\hat{s} = \Lambda_{zk}^{-1/2} E_{zk}' \Sigma_{vv}^{-1/2} Q_v P Y \quad (27)$$

Seja W_{vv} a matriz diagonal cujos elementos diagonais são os elementos da diagonal de Q_v , então

$$\Sigma_{vv}^{-1/2} = \frac{1}{\sqrt{\sigma^2}} W_{vv}^{-1/2} \quad (28)$$

que substituindo em (27), resulta

$$\hat{s} = \frac{1}{\sqrt{\sigma^2}} \Lambda_{zk}^{-1/2} E_{zk}' W_{vv}^{-1/2} Q_v P Y \quad (29)$$

ou ainda, fazendo

$$G = \frac{1}{\sqrt{\sigma^2}} P Q_v W_{vv}^{-1/2} E_{zk}' \Lambda_{zk}^{-1/2} \quad (30)$$

a (29) pode ser escrita na forma

$$\hat{s} = G' Y \quad (31)$$

Na presença de *outliers*, o vetor \hat{s} poderá ser decomposto segundo duas componentes

$$\hat{s} = s_o + \delta s = G' (Y + \delta Y) \quad (32)$$

sendo

$$\delta s = G' \delta Y \quad (33)$$

a componente que contém o efeito do erro e

$$s_o = G' Y \quad (34)$$

a componente de s livre do efeito do erro. Evidentemente que na ausência de erros grosseiros tem-se $\delta Y = 0$ e, portanto, $\delta s = 0$, ficando então $\hat{s} = s_o$.

Sem a presença de *outliers* a estatística

$$\chi^2 = \sum_{i=1}^k \hat{s}_i^2 \quad (35)$$

tem distribuição qui-quadrado com $\nu = k - 1$ graus de liberdade e parâmetros

$$E(\chi^2) = \nu \quad e \quad V(\chi^2) = 2\nu \quad (36)$$

Existindo *outliers*, a distribuição da estatística χ^2 é não-centrada (Aduol, 1987; Koch, 1985) com parâmetros

$$E(\chi'^2) = \nu + \tau_k \quad e \quad V(\chi'^2) = 2(\nu + \tau_k) \quad (37)$$

onde τ_k representa o parâmetro de não-centralidade da distribuição.

Portanto, para um certo nível de significância α , o teste para verificar a existência ou não de *outliers* pode ser formulado através das hipóteses

$H_0 : \tau_k = 0$ (ou seja, a presença de *outliers* não é significativa)

$H_1 : \tau_k \neq 0$ (ou seja, a presença de *outliers* é significativa).

Sendo constatada a presença de *outliers*, ou seja, se

$$\chi^2 = \sum_{i=1}^k \hat{s}_i^2 > \chi_{\alpha}^2(k) \quad (38)$$

tem-se que grandes valores para a estatística χ^2 implica em grandes valores de $|\hat{s}_i|$. Portanto, para o $|\hat{s}_i|_{\max}$ pode-se identificar na i -ésima coluna da matriz G , os elementos que correspondem aos coeficientes de \hat{s}_i . A aplicação do teste χ^2 é seqüencial, ou seja, constatada a existência de *outlier*, elimina-se o $|\hat{s}_i|_{\max}$, repetindo-se o processo consecutivamente até a estatística χ^2 não indicar mais a presença de *outliers*.

Considerando a matriz

$$G = \begin{bmatrix} g_{11} & g_{21} & \dots & g_{i1} & \dots & g_{k1} \\ g_{12} & g_{22} & \dots & g_{i2} & \dots & g_{k2} \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ g_{1n} & g_{2n} & \dots & g_{in} & \dots & g_{kn} \end{bmatrix} \quad (39)$$

então, se o maior valor de $|\hat{s}_i|$ for o i -ésimo, tem-se que

$$\hat{s}_i = g_{i1} Y_1 + g_{i2} Y_2 + \dots + g_{in} Y_n \quad (40)$$

onde, separando-se os elementos g_{ij} da coluna i que apresentam maior magnitude, ou seja, todos aqueles elementos dessa coluna que superam os seus correspondentes nas demais colunas.

Para reduzir o número de observações separadas como suspeitas através da análise da matriz G , pode-se utilizar um teste de discordância de *outliers*, como é o caso do teste de Ferguson (Barnett & Lewis, 1984) que utiliza a estatística

$$T = \frac{1}{ns^4} \sum_{i=1}^n (x_i - \mu)^4 \quad (41)$$

onde: n = número de observações;
 s = desvio padrão da amostra;
 x_i = i -ésima observação;
 μ = média conhecida da variável.

Os valores críticos para T , podem ser encontrados na referência (Barnett & Lewis, 1984). Esse teste é utilizado para um ou mais *outliers*, sendo que na presença de mais de um *outlier* o mesmo deve ser aplicado consecutivamente.

Com a redução do número de observações suspeitas, o próximo passo é identificar os *outliers*. Para isso deve-se estimar o vício de cada observação, utilizando-se de um novo ajustamento, introduzindo um novo vetor de parâmetros no modelo (1), ou seja, o modelo é colocado na forma,

$$\underline{Y} = A \underline{X} + B \underline{W} + \underline{V} \quad (42)$$

onde: \underline{W} = vetor $r \times 1$ dos r parâmetros (r vícios que serão estimados);

B = matriz $n \times r$ adequada ao modelo, ou seja, será uma matriz contendo tantas colunas quanto for o número de vícios a ser estimado, sendo cada coluna formada por $n-1$ elementos nulos e apenas um elemento igual a 1. A posição de cada elemento unitário deve estar em correspondência com a posição da observação que se quer estimar o vício.

O ajustamento do modelo (42) pelo método dos mínimos quadrados vai fornecer as estimativas dos vícios, bem como as estimativas dos erros padrões dos mesmos. Mas o simples fato de um vício estimado parecer grande não é suficiente para decidir se a observação correspondente é um *outlier*. Para isso é necessário sujeitar esses vícios estimados a um teste de hipótese. Os vícios estimados apresentam média zero, tendo distribuição normal com um certo desvio padrão desconhecido; então, para decidir se um certo vício estimado pode ser considerado significativo basta padronizá-lo e em seguida aplicar um teste de hipótese envolvendo a distribuição t de Student (Marques, 1982).

APLICAÇÃO EM FOTOTRIANGULAÇÃO

Para testar a metodologia descrita, realizou-se uma fototriangulação, com a simulação de um bloco composto de seis fotografias, consistindo em 34 pontos aerotriangulados e 150 pontos imagens.

As observações foram simuladas, considerando-se que os erros das observações (fotocoordenadas) apresentam comportamento normal com desvio padrão de 0,004 mm. Com base na simulação realizada as fotocoordenadas foram consideradas isentas dos erros sistemáticos.

Como modelo matemático funcional utilizou-se as Equações de Colinearidade, e o ajustamento simultâneo realizado através do método por feixe de raios (*bundle method*). Para o exemplo proposto, o ajustamento envolveu:

nº de parâmetros = $6 \times 6 + 34 \times 3 = 138$;

nº de observações = $150 \times 2 = 300$.

Não foram aplicadas injunções iniciais, optando-se dessa forma pelo ajustamento livre, utilizando como inversa da matriz das equações normais a pseudo-inversa (Moura, 1981). Portanto, utilizando o ajustamento livre, o número de graus de liberdade para o exemplo proposto é

$$k = n - u + d = 300 - 138 + 7 = 169$$

onde: k = nº de graus de liberdade;
 n = nº de observações;
 u = nº de parâmetros;
 d = deficiência de *rank*.

O modelo matemático funcional adotado, ou seja, as Equações de Colinearidade (Lugnani, 1987; Merchant, 1979) são

$$F(x) = x - x_0 - c \cdot \frac{(X - X_0)m_{11} + (Y - Y_0)m_{12} + (Z - Z_0)m_{13}}{(X - X_0)m_{31} + (Y - Y_0)m_{32} + (Z - Z_0)m_{33}} = 0$$

$$F(y) = y - y_0 - c \cdot \frac{(X - X_0)m_{21} + (Y - Y_0)m_{22} + (Z - Z_0)m_{23}}{(X - X_0)m_{31} + (Y - Y_0)m_{32} + (Z - Z_0)m_{33}} = 0 \quad (43)$$

Linearizando o modelo (43), obtém-se

$$A \underline{X} + \underline{L} = \underline{V} \quad (44)$$

onde: $A = \frac{\delta F}{\delta X_a}$ representa a matriz dos coeficientes (300×138);

\underline{X} = vetor das correções dos parâmetros (138×1);

$\underline{L} = \underline{L}_0 - \underline{L}_b$ (300×1), com $\underline{L}_0 = F(\underline{X}_0)$ sendo \underline{X}_0 o vetor dos valores aproximados dos parâmetros e \underline{L}_b o vetor das observações;

\underline{V} = vetor dos resíduos (300×1).

A solução do sistema (44) foi obtida através do ajustamento pelo método dos mínimos quadrados, utilizando-se de procedimento iterativo;

$$\underline{X} = -(A' P A)^{-1} A' P \underline{L} \quad (45)$$

ou

$$\underline{X} = -N^+ U \quad (46)$$

onde P representa a matriz diagonal dos pesos (300×300) e N^+ a pseudo-inversa da matriz das equações normais.

A Tabela 1 mostra resumidamente as observações simuladas e os correspondentes números das fotos e dos pontos.

Foto n°	Ponto n°	Observação n°	Fotocoordenada observada (em mm)
1	18	1	$x_1 = 105,99300$
1	18	2	$y_1 = 54,62100$
.	.	.	.
.	.	.	.
4	10	158	$x_{158} = 75,65700$
4	10	159	$y_{159} = 63,92600$
.	.	.	.
.	.	.	.
5	21	243	$x_{122} = -84,79000$
5	21	244	$y_{123} = -47,66700$
.	.	.	.
.	.	.	.
5	29	251	$x_{126} = -11,98000$
5	29	252	$y_{126} = -80,92600$
.	.	.	.
.	.	.	.
6	31	299	$x_{150} = -101,60400$
6	31	300	$y_{150} = -104,87000$

Tabela 1 - Observações simuladas.

Com a utilização de um processo aleatório, foram selecionadas 3 observações com o objetivo de contaminá-las com erros de magnitude entre -30σ e $+30\sigma$, sendo as magnitudes também selecionadas por um processo aleatório. Essa simulação resultou na seleção das seguintes observações e dos correspondentes erros (Tab.2).

N° da observação	Erro (em mm)	Observação (em mm)	Observação + erro (em mm)
159	-0,06578	63,92600	63,86022
243	-0,08862	-84,79000	-84,87862
252	0,03948	-80,92600	-80,88652

Tabela 2 - Observações contaminadas por erros.

Efetuada o ajustamento livre (Moura, 1981), determinou-se então o vetor das componentes principais padronizadas para o modelo linearizado (44), sendo que para esse modelo o vetor (31) é determinado por

$$\hat{s} = \frac{1}{\sqrt{\sigma^2}} \Lambda_{zk}^{-1/2} E_{zk}' W_{vv}^{-1/2} \hat{V} \quad (47)$$

$$\hat{s} = G' \hat{V}$$

onde: \hat{s} (169x1), G (300x169), \hat{V} (300x1). Como se pode notar o vetor das componentes principais padronizadas estimadas é constituído de 169 componentes, ou seja, igual ao número de graus de liberdade do modelo, correspondendo aos autovalores não nulos.

Determinado o vetor \hat{s} , calculou-se a estatística χ^2 , ou seja

$$\chi^2 = \sum_{i=1}^{169} \hat{s}_i^2 = 655,5 \quad (48)$$

Para um nível de significância de 5%, determinou-se o valor crítico

$$\chi^2_{0,05} (168) = 199,2 \quad (49)$$

Como $\chi^2 > \chi^2_{0,05} (168)$, rejeita-se a hipótese $H_0: \tau_k = 0$, ou seja, a presença de erro em uma ou mais observações (*outliers*) é significativa ao nível de significância de 5%.

No próximo passo, realizou-se uma análise criteriosa sobre a matriz G , como descrita em Fundamentos Teóricos do Método, para selecionar os elementos g_{ij} suspeitos de carregarem as observações contaminadas. Foram selecionados 35 elementos desta matriz cujas posições correspondem às observações de números:

5	7	8	31	32	39	42	50	54
59	79	88	91	92	117	119	120	129
159	161	171	177	179	182	201	203	207
210	211	241	243	252	264	267	276	

Para reduzir o número de observações suspeitas, aplicou-se o teste de discordância de *outliers* de Ferguson, o qual utiliza a estatística T (41). A estatística T de Ferguson mostrou-se significativa para 5 observações, ao nível de significância de 5% como mostra a Tabela 3.

Estatística T(n) (n=n° de obs.)	Valor Crítico (nível = 5%)	N° da observação selecionada
T(35) = 7,89	4,2	243
T(34) = 5,17	4,2	159
T(33) = 4,92	4,2	252
T(32) = 4,50	4,2	117
T(31) = 4,63	4,2	8
T(30) = 3,81	4,2	não é significativo

Tabela 3 - Observações selecionadas pelo teste de Ferguson.

Na última etapa procedeu-se à identificação dos *outliers* entre as cinco observações selecionadas. Para isso, inicialmente foi necessário estimar os vícios de cada uma delas. A estimação foi feita através de um novo ajustamento onde o modelo matemático linearizado é do tipo,

$$AX + BW + L = V \quad (50)$$

onde: B = matriz 300x5 dos coeficientes dos vícios;
 W = vetor 5x1 dos vícios.

O vetor \hat{W} estimado será um vetor de parâmetros (vícios) do tipo,

$$\hat{W} = \begin{bmatrix} \hat{w}_8 \\ \hat{w}_{117} \\ \hat{w}_{159} \\ \hat{w}_{243} \\ \hat{w}_{252} \end{bmatrix} \quad (51)$$

Realizado o ajustamento, os vícios estimados e respectivos erros padrões encontram-se na Tabela 4.

Nº da observação	Vício - \hat{w}_i	Erro padrão
8	0,00309	0,00540
117	0,02193	0,02174
159	0,05206	0,01521
243	0,08879	0,00569
252	- 0,04049	0,00494

Tabela 4 - Vícios e erros padrões estimados.

Finalmente, calculou-se a estatística t para cada vício estimado que através de teste de hipótese descrito em Fundamentos Teóricos do Método, permitiu identificar os outliers.

A Tabela 5 mostra o valor da estatística t para cada uma das cinco observações selecionadas e as conclusões obtidas. O teste de hipótese foi aplicado para um nível de significância de 5%, sendo o número de graus de liberdade dado por

$$v = n - u - r + d = 164 \quad (52)$$

onde r = número de vícios estimados.

O valor crítico de t, para o teste bilateral, é dado por

$$t_{0,025}(164) = 1,97$$

Nº da observação	Estatística t	Conclusão
8	0,57	não é outlier
117	1,01	não é outlier
159	3,42	é outlier
243	15,59	é outlier
252	-8,20	é outlier

Tabela 5 - Valores da estatística t.

Como pode-se observar na Tabela 5, as únicas observações identificadas como outliers são as de números: 159, 243 e 252, ou seja, exatamente aquelas que foram contaminadas.

Os vícios finais estimados dessas três observações foram obtidos efetuando-se um novo ajustamento, utilizando um modelo do tipo (50), substituindo-se apenas o vetor \hat{W} : 5X1, por outro de dimensão 3x1, ou seja,

$$\hat{W}_1 = \begin{bmatrix} \hat{w}_{159} \\ \hat{w}_{243} \\ \hat{w}_{252} \end{bmatrix} \quad (53)$$

e a matriz B:300x5 por outra matriz adequada B₁: 300x3.

O novo ajustamento, para o presente caso, forneceu os vícios estimados finais como mostra a Tabela 6.

Nº da observação	Vício estimado
159	0,06625
243	0,08997
252	- 0,04043

Tabela 6 - Vícios estimados finais.

Com os vícios estimados, pode-se ainda corrigir as observações contaminadas e testar novamente se a existência de outliers é significativa. Essa correção foi feita e calculou-se novamente a estatística χ^2 obtendo-se

$$\chi^2 = \sum_{i=1}^{169} \hat{s}_i^2 = 164,1$$

Portanto,

$$\chi^2 < \chi^2_{0,05}(168) = 199,2$$

ou seja, o resultado da estatística não é mais significativo; portanto, as observações não contém outliers.

CONSIDERAÇÕES FINAIS

O uso mais freqüente de técnicas estatísticas multivariadas só tem sido possível graças à evolução dos computadores, principalmente no que diz respeito ao aumento de velocidade de processamento e da capacidade de armazenar informações. O que se pretendeu com esse trabalho foi mostrar que as técnicas multivariadas podem ser bastante úteis na detecção e identificação de múltiplos outliers em fototriangulação. Os resultados do experimento realizado mostram o potencial e eficiência do método da Análise de Componentes Principais, na detecção e identificação de erros observacionais em Ajustamento. Como sugestão para estudos posteriores seria interessante pesquisar o tempo de processamento do método multivariado quando comparado com os métodos usuais e também a comparação de resultados para diferentes configurações em fototriangulação.

REFERÊNCIAS BIBLIOGRÁFICAS

Aduol, F. W. O. 1987. Detection of outliers in geodetic networks using principal component analysis and bias parameter estimation. *Technical Report*, 2: 1-104.

Berberan, A. 1992. Outlier detection and heterogeneous observations: a simulation case study. *Australian Journal of Geodesy, Photogrammetry and Surveying*, 56: 49-61.

Barnett, V. & Lewis, T. 1984. *Outliers in statistical data*. New York, John Wiley, 463p.

Clerici, E. & Harris, M.W. 1980. A premium-protection method applied to detection and rejection of erroneous observations. *Manuscripta Geodaetica*, 5: 283-298.

Crosilla, F. 1983. Procrustean transformation as a tool for the construction of a criterion matrix for control net works. *Manuscripta Geodaetica*, 8: 343-370.

Cross, P.A. & Price, D.R. 1985. A strategy for the distinction between single and multiple gross errors in geodetic networks. *Manuscripta Geodaetica*, 10: 172-178.

- Dalmolin, Q. & Marques, J. M. 1993. O uso da análise multivariada na detecção de outlier em fototriangulação. *Anais, 16º Congresso Brasileiro de Cartografia*, Rio de Janeiro. SBC, 2: 425-429.
- Davies, L. & Gather, U. 1993. The identification of multiple outliers. *Journal of the American Statistical Association*, 423 (88): 782-792.
- El-Hakim, S.F. 1981. A practical study of gross-error detection in bundle adjustment. *The Canadian Surveyor*, 35 (4): 373-386.
- Firkowski, H. 1986. *Confiabilidade da fototriangulação: configuração e detecção de erros no apoio de campo*. Curso de Pós-Graduação em Ciências Geodésicas, Universidade Federal do Paraná. Dissertação de mestrado. 245 p.
- Jackson, J. E. & Mudholkar, G. S. 1979. Control procedure for residuals associated with principal component analysis. *Technometrics*, 3 (21): 341-349.
- Johnson, R.A. & Wichern, D.W. 1988. *Applied multivariate statistical analysis* Prentice-Hall. Englewood Cliffs. 607 p.
- Koch, K.R. 1985 *Statistical analysis of geodetic data with applications for the deformation analysis*. Curitiba, Universidade Federal do Paraná, 125p.
- Kubik, K. & Wang, Y. 1991. Comparison of different principles for outlier detection. *Australian Journal of Geodesy, Photogrammetry and Surveying*, 54: 67-80.
- Lugnani, J.B. 1987. *Introdução à fototriangulação*. Curitiba, Universidade Federal do Paraná. 134p.
- Marques, J.M. 1982. *Análise estatística das discrepâncias parciais de uma rede de nivelamento de 1ª ordem*. Curso de Pós-Graduação em Ciências Geodésicas, Universidade Federal do Paraná. Dissertação de mestrado. 153p.
- Merchant, D.C. 1979. *Analytical photogrammetry: theory and practice*. The Ohio State University, Ohio 2v.
- Mitshita, E.A. 1986 *Deteção de erros grosseiros nas aerotriangulações*. Curso de Pós-Graduação em Ciências Geodésicas, Universidade Federal do Paraná. Dissertação de mestrado. 246p.
- Morrison, D.F. 1976. *Multivariate statistical methods*. McGrawHill, Tokyo. 415p.
- Moura, J.O. 1981. *Aplicação do ajustamento livre à fotogrametria*. Curso de Pós-Graduação em Ciências Geodésicas, Universidade Federal do Paraná. Dissertação de mestrado. 122p.
- Ou, Z.Q. 1989. Sequential tests for outliers in the general linear model. *Australian Journal of Geodesy, Photogrammetry and Surveying*, 50: 37-49.
- Patterson, R.C. 1985. Outlier detection: some distribution theory. *Australian Journal of Geodesy, Photogrammetry and Surveying*, 43: 1-30.
- Pope, A.J. 1976. The statistics of residuals and the detection of outliers. *NOAA Technical Report NOS 65 NGS 1* : 1-133.
- Silva, J.F.C. 1987. *A inferência bayesiana e a detecção e localização automática de erros grosseiros em fototriangulação por feixe de raios*. Curso de Pós-Graduação em Ciências Geodésicas. Universidade Federal do Paraná. Tese de doutorado. 88p.