

Tratamento de Outliers em Redes Topográficas Locais Obtidas por GPS

M.Sc. Antonio José Prado Martins Santos ¹
Prof. Dr. Ricardo Ernesto Schaal ²

EESC-USP- Depto. De Eng. De Transportes
CEP 13566-970 São Carlos SP

¹ aj_pms@yahoo.com.br

² schaal@sc.usp.br

Resumo: Em levantamentos por meio do GPS de redes topográficas se faz necessário conhecer a qualidade das coordenadas estimadas de acordo com o tipo de aplicação a que se destinam. Este trabalho aborda as teorias de análise de qualidade destas redes, baseando-se nas teorias de confiabilidade propostos por Baarda em 1968. Conjuntamente propõem dois modos estratégicos para o tratamento dos dados com presença de outliers. Os resultados dos ajustamentos por meio das estratégias foram comparados com os do programa comercial Ski-Pro e também validados por medidas de campo feitas com Estação Total.

Palavras chaves: redes topográficas, *outliers*, confiabilidade.

Abstract: In land surveying by means of GPS network it is necessary to know the quality of the estimated coordinates with the class of final application. This article goes over the theory on the analyses of network quality, based on the reliability theory proposed by Baarda in 1968. Simultaneously it is proposed two strategies modes for handling data with outliers. The results from these strategies were compared with results obtained with the Ski-Pro commercial software together with a field validation using a Total Station.

Key words: Topographic networks, outliers, reliability

1 - Introdução

O conjunto de marcos materializados no terreno e referenciados a um sistema de coordenadas topográficas horizontais e/ou verticais forma a chamada redes topográficas locais, que serve de base para trabalhos como: atualização e elaboração de plantas cadastrais, referência a todos os serviços topográficos de demarcação, de anteprojeto, implantação e acompanhamento de obras de engenharia (torres, pontes, barragens, edificações, túneis, extração de minérios), cadastro imobiliário, registros públicos, multifinalitários etc.

As medições feitas nestes marcos, por meio de uma rede de vetores GPS, permitem a determinação de suas posições mediante um ajustamento pelo Método dos Mínimos Quadrados, porém erros grosseiros ou *outliers* podem estar presentes nas medidas de forma a influenciar o resultado final do ajustamento de uma rede.

Uma medida com *outlier* apresenta-se de forma distinta em relação às outras e é considerada como observações inconsistentes.

O estudo de *outliers* é de fundamental importância para trabalhos que requerem análise estatística dos dados observados. Todo pesquisador que faz medidas em seus experimentos já se deparou com observações que se distanciam das demais.

Segundo Leick (2004) a eficiência dos testes estatísticos para erros grosseiros está associada ao grau de confiabilidade da rede, ou seja, quanto mais confiável for a rede, maior será a probabilidade de se detectar os erros através dos testes estatísticos.

Um teste aplicado para a detecção de possíveis *outliers* é o teste global *qui-quadrado*, χ^2 , utilizado com um alfa unidimensional, α_0 , obtido pelo monograma de Baarda (1968) em função:

a) do alfa multidimensional, α ;

b) da potência do teste, $1 - \beta_0$;

c) do grau de liberdade, S , ou traço da matriz de redundância R , (Ver em Kuang, 1996).

É dito possíveis *outliers* devido ao teste global ser um indicativo de inconsistências no ajustamento, ou seja, possibilidade de existência de:

1- superestimação ou subestimação da matriz peso;

2- presença de *outliers*;

3- alguma inconsistência do modelo matemático com as observações.

No caso particular de ciências geodésicas, Vanicek & Krakiwsky (1986), afirmam que o modelo matemático relacionando os dados coletados para certos parâmetros desconhecidos é muito bem definido, descartando assim, esta possibilidade. Quando do caso de super ou subestimação da matriz peso, torna-se necessário realizar um escalonamento, ou seja, multiplicar a matriz variância covariância por um fator de escala geralmente igual à variância a posteriori encontrada.

A estatística do teste para a i -ésima observação é dada pelo teste *data snooping* de Baarda (Baarda, 1968), tendo como principal objetivo a localização de possíveis erros embutidos nas observações. No entanto, os testes global e *data snooping* requerem o conhecimento da precisão das observações. Existem casos em que se tem um conjunto de observações para ajustar, e que a precisão das observações não é conhecida. Nestes casos, após o ajustamento não é realizado o teste global nem o *data snooping*, mas sim o teste *tau*, que é uma adaptação do teste *data snooping* com vistas a se adequar a tal situação (Kuang, 1996).

Os testes globais, *data snooping* e *tau* são bons para as análises pós-ajustamento, detectando e localizando os possíveis erros, porém estes testes não indicam se os erros detectados e localizados são significativamente influentes nos resultados da rede. Tais erros ocasionam a alteração dos resultados, alterando as coordenadas dos pontos da rede. Desta forma, torna-se necessário a utilização dos testes de confiabilidade para indicar o quanto as observações são confiáveis.

Sendo detectado erros como significativamente influentes após a realização dos testes de confiabilidade, o ideal seria voltar a campo e medir novamente os vetores que apresentaram componentes com erros, porém na maioria dos casos torna-se inviável por questão de tempo de trabalho e custos envolvidos neste procedimento. Sendo assim, uma solução seria a eliminação do(s) erro(s) existente(s).

2 – Medidas de confiabilidade

2.1-Confiabilidade interna

A confiabilidade interna quantifica a menor porção do erro existente na observação que pode ser localizado com uma dada probabilidade, ou seja, indica o erro mínimo que se encontra em uma observação que é sensível ao teste (Kavouras, 1982; Kuang, 1996).

A equação que descreve este erro mínimo detectável, ∇_{oli} , é dada por:

$$\nabla_{oli} = \sigma_{li} \cdot \frac{\delta_o}{\sqrt{r_i}} \quad (1)$$

onde, σ_{li} é a precisão da i-ésima observação; δ_0 é o parâmetro de não-centralidade do teste unidimensional (ver Kuang, 1996) e r_i a redundância parcial da i-ésima observação obtida na diagonal da matriz de redundância R , dada por:

$$R = Q_{vv}P = I - A(A^T P A)^{-1} A^T P \quad (2)$$

onde, Q_{vv} é a matriz cofatora dos resíduos; P é a matriz peso; I é a matriz identidade e A é a matriz coeficiente dos parâmetros da equação das observações.

Em seguida estima-se o possível erro embutido nas observações, pela expressão:

$$\nabla l_i = - \frac{\hat{v}_i}{r_i}, \quad (3)$$

onde, $\hat{v}_i$ é o resíduo da i-ésima observação.

Um erro grosseiro será significativamente influente se o erro presente na medida for maior que o erro mínimo detectável, (Kuang, 1996).

$$\nabla l_i > \nabla_0 l_i \quad (4)$$

Sendo detectado erros como significativamente influentes, no caso de redes obtidas com GPS podemos fazer a eliminação dos *outliers* das seguintes formas:

✓ **Estratégia 1 – Eliminação das componentes dos vetores que apresentam erros grosseiros:**

Para a realização desta etapa, se retira dos dados iniciais de ajustamento as componentes $\Delta X, \Delta Y$ e/ou ΔZ , de cada vetor, que venham a ser detectados como erros grosseiros e novo ajustamento é realizado. O mesmo processo se repete caso que novos *outliers* venham a ser detectados. Cabe observar que os programas comerciais existentes não possibilitam a eliminação da componente com presença de erros, mas sim a eliminação de todo o vetor (abordado na estratégia dois).

✓ **Estratégia 2 – Eliminação completada dos vetores com erros grosseiros:**

Da mesma forma que a eliminação de componentes, elimina-se os vetores que apresentarem componentes com erros e novos ajustamentos serão realizados. O processo se repete até a eliminação total dos erros.

2.2 - Confiabilidade Externa

A confiabilidade externa é definida como a influência de cada um dos erros grosseiros detectáveis, ∇l_i , nos parâmetros do ajustamento ou nas funções dos parâmetros. Segundo Leick (2004), a confiabilidade interna consistente de uma rede não garante automaticamente coordenadas confiáveis e para isso o estudo necessita ser complementado com a análise da confiabilidade externa, a qual irá estimar o efeito dos erros nos parâmetros finais.

A estimativa dos parâmetros com presença de erro grosseiro é dada por:

$$\hat{X} = - N^{-1} A^T P (L_b - e_i \nabla l_i) \quad (5)$$

A influência do erro ∇l_i sobre as coordenadas estimadas é dada pela equação:

$$\nabla_X = N^{-1} A^T P e_i \nabla l_i \quad (6)$$

Onde:

N , matriz dos coeficientes das equações normais;

A , Matriz das derivadas das equações de observação em relação às incógnitas;

P , Matriz dos pesos das observações;

L , Vetor da diferença entre o vetor dos valores calculados e o vetor dos valores observados;

e_i , i-ésima coluna de uma matriz identidade $n \times n$.

O impacto de um erro marginal detectável $\nabla_0 l_i$ é:

$$\nabla_{X_{oi}} = N^{-1} A^T P e_i \nabla_0 l_i \quad (7)$$

Detalhes do estudo de confiabilidade também podem ser vistos em Amorim (2004); Leick (2004); Kuang (1996); Kavouras (1982).

3 – Resultados Experimentais

Foi planejada a implantação no Campus II da USP de São Carlos de uma rede constituída por 13 vértices.

A distribuição dos pontos foi planejada buscando observar fatores como: posição dos pontos da rede em relação à acessibilidade, proximidade de áreas construídas e regiões com tendência de crescimento (construções futuras).

A figura 1 apresenta a rede planejada e as distâncias horizontais entre os vértices, obtidas por meio do programa para desenhos, Autocad, versão 2000.

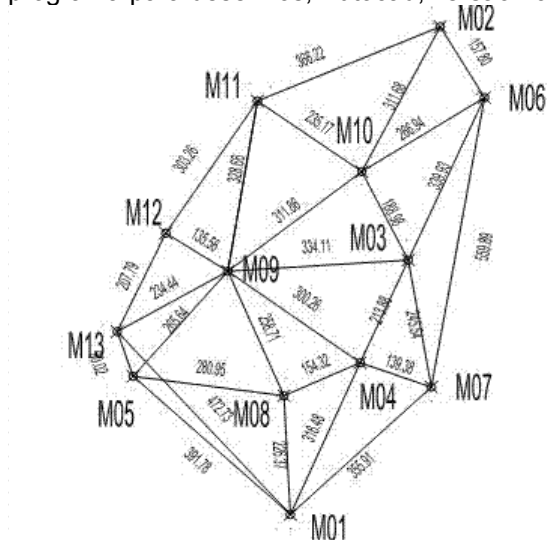


Figura 1: Projeto da rede do Campus II – unidade em (m)

em cada estação de aproximadamente 30 minutos.

Esta rede foi implantada objetivando: Marcos com coordenadas confiáveis de referência para as futuras construções no campus; levantamentos cadastrais em sua área de abrangência com o uso de Estação Total e/ou GPS. Os marcos foram construídos diretamente no solo por meio de estruturas de concreto (figura 2).

Para o posicionamento das estações desta rede, foi utilizada uma estação base, a M02 (ver figura 1), advinda da Rede Estadual do Estado de São Paulo (RESP) por transporte de coordenadas. A técnica de posicionamento utilizada foi a estática, empregando-se dois receptores GPS de frequência L1 (Leica 9400). Utilizou-se uma taxa de coleta de dados de 10 segundos, ângulo de elevação mínimo de 15 graus e período de ocupação



Figura 2: modelo dos marcos implantado no Campus II

3.1 - Processamento e Ajustamento dos dados

Os dados coletados foram pós-processados utilizando-se o programa Ski-Pro 2.1 (Leica). Durante a fase de processamento das linhas de base, retirou-se os satélites que apresentavam resíduos altos e todas as linhas de base tiveram as ambigüidades fixadas.

O ajustamento da rede foi realizado primeiramente no próprio Ski-Pro e em seguida foi feito um ajustamento por uma rotina desenvolvida em MATHCAD 2001 Professional onde foi levado em consideração as duas estratégias proposta anteriormente.

Para o ajustamento da rede levaram-se em consideração os seguintes parâmetros e testes estatísticos: Nível de confiança de 68%; nível de significância unidimensional α_0 de 5%; nível de

significância bidimensional ($\alpha_0 = 59\%$) obtido pelo monograma de Baarda (1968); potência do teste β_0 de 80%; análise do teste F (Ski-Pro) e o teste global *qui-quadrado* (MATHCAD); teste *Data snooping W* e análise da confiabilidade interna e externa da rede.

No ajustamento feito no MATHCAD observou-se que o teste global foi rejeitado, indicando que algum problema ou erro existe no ajustamento ou nos dados observados.

Para que o teste *qui-quadrado* fosse aceito (tabela 1), realizou-se um escalonamento da matriz variância-covariância por um valor aproximadamente igual à variância a posteriori encontrada no ajustamento.

Tabela 1 - Teste *qui-quadrado* aceito após escalonamento da MVC para ($\alpha_0 = 59\%$)

Variâncias		Intervalo de aceitação do teste X^2 bidimen.		X^2 Unidimen.		Result.
Priori	Posteriori	Inferior	Superior	Crítico	Calcu.	
σ_0^2	$\hat{\sigma}_0^2$	$X^2_{\left(S, \frac{\alpha}{2}\right)}$	$X^2_{\left(S, 1 - \frac{\alpha}{2}\right)}$	$X^2_{(S, 1-\epsilon)}$	χ^{*2}	
1	0,92	65,16	77,89	68,53	65,97	Aceito

Passando no teste *qui-quadrado* foi realizado o teste *Data snooping* de Baarda (1968) e os seguintes erros foram localizados (tabela 2):

Tabela 2 - Erros detectados pelo teste *Data snooping*:

Vetor	Componente do Vetor	(w) Estimado	(w) Crítico
M01-M05	ΔY	2,05	- $1.96 \leq W_{EST} \leq 1.96$
M06-M10	ΔY	-2,96	
M03-M10	ΔY	2,30	

A tabela 3 mostra a análise de significância dos erros localizados calculado pela equação (4).

Tabela 3 - Significância dos erros

		$\alpha_0 = 5\%, \beta_0 = 20\%, \delta_0 = 2.8$								
Est. Ref.	Relat.	$(\nabla_0 l_i)$ $\Delta X(m)$	$(\nabla_0 l_i)$ $\Delta Y(m)$	$(\nabla_0 l_i)$ $\Delta Z(m)$	(∇l_i) $\Delta X(m)$	(∇l_i) $\Delta Y(m)$	(∇l_i) $\Delta Z(m)$	Aceitação das observações		
M01	M05	16,38	15,76	10,13	4,36	12,89	6,79	SIM	SIM	SIM
M06	M10	16,47	18,36	11,85	9,24	21,65	2,86	SIM	NÃO	SIM
M03	M10	20,97	24,08	10,94	7,77	22,11	0,46	SIM	SIM	SIM

Pela tabela 3, dos erros grosseiros detectados pelo teste *Data snooping* apenas a componente Y do vetor M06>M10, seria significativamente influente no resultado de confiabilidade da rede. Neste trabalho, foi levado em consideração que todos os erros detectados (tabela 2) seriam influentes.

3.2 - Resultados das estratégias aplicadas:

Estratégia 1 – Depois de eliminadas as componentes dos vetores que possuíam *outliers* (tabela 2), foi realizado novo ajustamento na rede. Mais uma vez, o teste χ^2 para o alfa multidimensional de Baarda (1968), que é de 59%, foi rejeitado, necessitando, para ser aceito, de um escalonamento da matriz variância-covariância. Foi aplicada constante igual a 8, denominada de fator de escala.

Ao realizar mais uma vez a detecção de outliers pelo teste *Data snooping*, novos *outliers* surgiram. O mesmo processo de eliminação e ajustamento foi repetido até a eliminação total das componentes com presença de outliers.

Estratégia 2 - A eliminação dos vetores que possuíam *outliers* passou por três novos ajustamentos, sendo que no final foram eliminados ao todo seis vetores. Houve uma redução de aproximadamente 25% no valor de redundância do sistema.

A tabela 4 mostra as coordenadas e precisões de três pontos escolhidos da rede, obtidas com o Ski-Pro e com as duas estratégias estudadas.

Tabelas 4 - Coordenadas e suas precisões

Marco	Ajustamen	Coord (X) (m)	σ_x (m)	Coord (Y) (m)	σ_Y (m)	Altura(H) (m)	σ_H (m)
M01	Sky-Pro						
	Estratégia 1	149723.1064	0,002	249071.3156	0,002	840.3537	0,005
		149723.1061	0,003	249071.3153	0,004	840.3508	0,004
	Estratégia 2	149723.1069	0,003	249071.3162	0,004	840.3562	0,004
M06	Sky-Pro						
	Estratégia 1	150081.7953	0,003	249865.0521	0,002	838.2767	0,005
		150081.7945	0,002	249865.0508	0,004	838.2700	0,004
	Estratégia 2	150081.7973	0,002	249865.0511	0,004	838.2682	0,004
M09	Sky-Pro						
	Estratégia 1	149607.5777	0,001	249534.4079	0,002	846.3680	0,004
		149607.5748	0,002	249534.4070	0,003	846.3592	0,003
	Estratégia 2	149607.5785	0,003	249534.4082	0,004	846.3623	0,004

Na tabela acima as maiores diferenças planimétricas são de aproximadamente 4 mm o que é irrelevante para uma rede para apoios de trabalhos topográficos.

Cabe observar que diferentes programas de ajustamento podem apresentar diferentes formas para a construção de sua matriz peso, bem como, distintos valores para a atribuição do fator de escalonamento da matriz variância-covariância.

A figura 3 mostra a confiabilidade externa obtida pela equação (6). As estratégias 1 e 2 apresentaram bons resultados. Isto já era esperado uma vez que todos os erros grosseiros identificados foram eliminados. Ainda assim, nota-se que a estratégia 1 (somente eliminação de componentes) apresentou melhores resultados que a estratégia 2.

Analisando os dados de confiabilidade externa do Ski-Pro (figura 3) as influências dos erros foram homogêneas, porém, houve uma maior influencia em todos os pontos da rede em relação às estratégias 1, 2. Isto ocorreu devido ao programa Ski-Pro ter detectado maior número de erros que as estratégias aplicadas e não os ter eliminado.

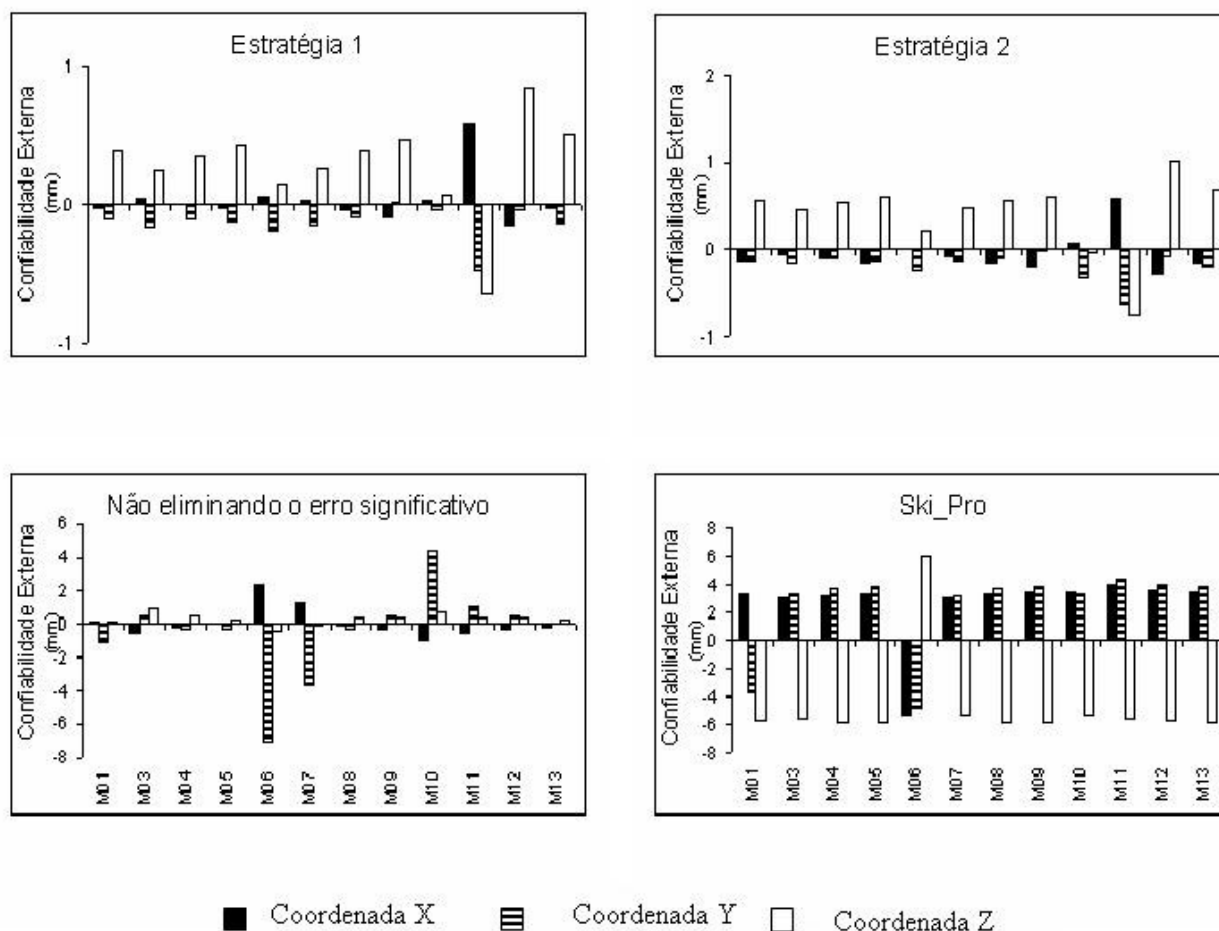


Figura 3: Confiabilidade externa no Ski-Pro e nas estratégias

O gráfico da figura 3 que não considerou a eliminação dos erros detectados na rotina em MATHCAD mostra a maior influência destes erros, principalmente o erro da componente Y do vetor *M06-M10* influenciando bastante nas coordenadas dos pontos *M06*, *M07* e *M10*.

Para testar a qualidade da rede foram realizadas medidas lineares e angulares com uma Estação Total Leica TC-403 com precisão linear de 2 mm + 2ppm e angular de 3 segundos.

Tabela 5 - Resíduos das distâncias entre estação e ré nas diferentes estratégias e E.T.

Vetor	Distâncias em (m) obtidas por coordenadas				Diferença Obtida (mm)		
	SKI	ESTR. 1	ESTR. 2	Est. Total			
M03-M02	448.24 5	448.24 6	448.24 5	448.24 5	0	1	0
M03-M06	339.92 5	339.92 5	339.92 6	339.92 5	0	0	1
M04-M03	214.61 8	214.61 9	214.61 7	214.61 8	0	1	1
M08-M04	154.31 2	154.31 2	154.31 2	154.31 2	0	0	0

Pelos resultados obtidos na tabela 5 observa-se que a variação entre a distância medida com Estação Total (E.T.) e as obtidas nas estratégias propostas ficou na ordem milimétrica. Sendo assim, a rede implantada tem uma boa amarração linear entre seus vértices.

Nas diversas estratégias foram obtidos bons resultados quanto à orientação, pois a diferença entre as coordenadas para um mesmo ponto obtido por caminhos distintos, por exemplo, o ponto PT01 (figura 4), ficou abaixo de 6 mm na coordenada X e de 6 mm na coordenada Y, ocorrendo em proporções semelhantes para todos os pontos levantados.

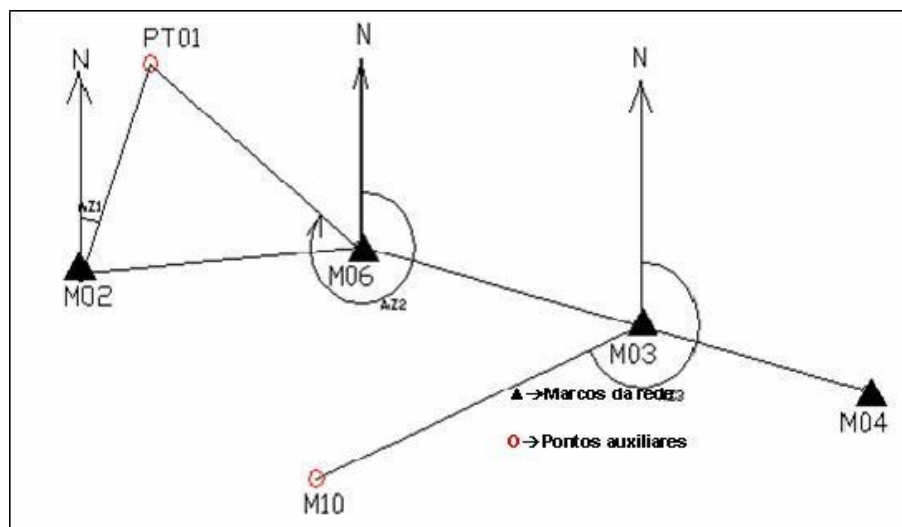


Figura 4: Controle angular da rede

4 - Conclusões

Aplicando a teoria da confiabilidade de Baarda constatou-se que em redes topográficas nem todos os erros, mesmo identificados como significativamente influentes, são prejudiciais no resultado final da rede.

Porém o usuário deve tomar bastante cuidado e de preferência fazer uso das técnicas aqui abordadas de forma a observar qual método de ajustamento fornecerá o melhor resultado.

Na necessidade de eliminação de *outliers* (quando se consideram influentes os erros detectados), a estratégia 1 (eliminação de componentes) forneceu melhores resultados devido a eliminar apenas as componentes com erros grosseiros, não reduzindo tanto o número de redundância do sistema, ocorrendo o contrário com a estratégia 2.

Pode-se afirmar que a rede ajustada nas diversas estratégias obteve bons resultados angulares e lineares e que um operador partindo de qualquer ponto da rede fechará com erros milimétricos em qualquer outro ponto da mesma.

Importante lembrar que esta é uma rede piloto e que está em um local privilegiado com poucas construções e com dimensões quilométricas. No caso de uma rede de maior envergadura e em local de maior densidade urbana se pode esperar divergência maior entre os ajustamentos, mas, a análise de qualidade sempre procede.

5 - Referências Bibliográficas

- Amorim, G. P.** *Confiabilidade de Rede GPS de Referência Cadastral Municipal, Estudo de Caso: Rede do Município de Vitória (ES)*. Dissertação (Mestrado). Escola de Engenharia de São Carlos, Universidade de São Paulo. São Carlos, 2004, 194 p.
- Baarda, W.** *A testing procedure for use in geodetic networks*. Netherlands Geodetic Commission. Publications on Geodesy. Série 2, N° 5. Delft . Netherlands, 1968.
- Gemael, C.** *Introdução ao ajustamento de observações: aplicações de Geodésia*. 1ª edição. Universidade Federal do Paraná. Curitiba, 1994
- IBGE - Fundação Instituto Brasileiro de Geografia e Estatística.** *Especificações e normas gerais para levantamentos geodésicos*. Resolução PR No 22. Rio de Janeiro, 1983.

- Kavouras, M.** *On detection of outliers and the determination of reliability in geodetic networks*. Technical report No 87. Department Surveying Engineering. University of New Brunswick Fredericton, Canada (1982).
- Kuang, S.** *Geodetic network analysis and optimal design: Concepts and Applications*. 1st ed. Michigan: Ann Arbor. Inc. Chelsea, Michigan, 1996.
- Leick, A.** *GPS: Satellite Surveying*. 2nd,ed. New York: John Wiley & Sons, Incorporation, 2004.
- Moraes, C. V.** *Análise de erros grosseiros e confiabilidade de redes geodésicas*. Revista Cartografia e Cadastro. Instituto Português de Cartografia e cadastro. Nº 8. Lisboa, 1998.
- Souza, G. C.** *Operacionalidade de Redes Geodésicas de Apoio ao Cadastro Rural*. Tese (Doutorado). Escola de Engenharia de São Carlos, Universidade de São Paulo. São Carlos, 2005.148 p.
- Santos, A. P. M. S.** *Proposta de ajustamento para melhoria da confiabilidade e precisão dos pontos de redes geodésicas para fins topográficos locais*. Dissertação (Mestrado). Escola de Engenharia de São Carlos, Universidade de São Paulo. São Carlos, 2006.165 p.
- Teixeira, N. N.; Ferreira, L. D. D.** (2003). *Análise da confiabilidade de redes geodésicas*. Artigo. Boletim de Ciências Geodésicas v.9, nº 2. Curitiba. Paraná, 2003. p 199-216.
- Vanicek, P.; Krakiwsky, E. J.** *Geodesy: The Concepts*. 2nd ed, Amsterdam. New York. Oxford: North_Holland Publishing Company, (1986).