

## ANALYTICAL QUALITY ASSESSMENT OF ITERATIVELY REWEIGHTED LEAST-SQUARES (IRLS) METHOD

*Avaliação da qualidade analítica do método dos mínimos quadrados iterativo com a atribuição de novos pesos*

JIANFENG GUO

Info Engineering University  
62 Kexuedadao Rd, P.O. Box 2201-160, Zhengzhou, 450001, China  
[jianfeng.guo@gmail.com](mailto:jianfeng.guo@gmail.com)

### ABSTRACT

The iteratively reweighted least-squares (IRLS) technique has been widely employed in geodetic and geophysical literature. The reliability measures are important diagnostic tools for inferring the strength of the model validation. An exact analytical method is adopted to obtain insights on how much iterative reweighting can affect the quality indicators. Theoretical analyses and numerical results show that, when the downweighting procedure is performed, (1) the precision, all kinds of dilution of precision (DOP) metrics and the minimal detectable bias (MDB) will become larger; (2) the variations of the bias-to-noise ratio (BNR) are involved, and (3) all these results coincide with those obtained by the first-order approximation method.

**Keywords:** IRLS; Outlier; DOP; MDB; BNR.

### RESUMO

A técnica de mínimos quadrados iterativa com reponderação dos pesos tem sido amplamente usada na literatura geodésica e geofísica. As medidas confiáveis são instrumentos diagnósticos importantes para inferir a força da validação do modelo. Um método analítico exato é adaptado para obter a compreensão do quanto a reponderação iterativa dos pesos pode afetar os indicadores de qualidade dos resultados. Análises teóricas e numéricas mostram que, quando o procedimento de diminuição dos pesos é feita: 1) a precisão métrica e a tendência mínima detectável, aumenta; 2) As tendências da variação média do ruído estarão envolvidas e; 3)

Todos os resultados coincidirão com aqueles obtidos com o método de aproximação de primeira ordem.

**Palavras-chave:** IRLS; Outlier; DOP; MDB; BNR.

## 1. INTRODUCTION

Least-squares (LS) method exhibits a poor performance in the presence of outliers. A reliable alternative to LS is given by the robust regression techniques. Iteratively updating the weights yields the iteratively reweighted least squares (IRLS) algorithm, which is the most common method for computing M-estimates (HUBER 1981, HUBER and RONCHETTI 2009). In fact, IRLS has been extensively employed in geodetic and geophysical literature (CHANG and GUO 2005, RANGELOVA et al. 2009, GUO et al. 2010, COLLILIEUX et al. 2012).

Under the assumption of only one outlier exists, Baarda (1968) developed his famous testing procedure in the framework of mean-shift outlier model (GUO 2013), which ultimately led to the reliability theory. Extension of reliability measures for correlated observations was discussed by Wang and Chen (1994), Schaffrin (1997) and Ou (1999). Generalized measures of reliability in the presence of multiple outliers were addressed by Knight et al. (2010).

There are two types of reliability measures: internal and external. Both of them are important diagnostic tools for inferring the strength of the model validation (cf. TEUNISSEN 1985, VERHAGEN 2002, LEICK 2004). By using the first-order approximation, Guo et al. (2011) investigated the variation characteristics of minimal detectable bias (MDB) and the bias-to-noise ratio (BNR) measures for an iterative robust M-estimator. This contribution serves a twofold purpose: (1) to evaluate the impact of iterative reweighting on the quality indicators by using an exact analytical method, and (2) to assess the adequacy of the first order approximation method.

## 2. ITERATIVELY REWEIGHTED LEAST-SQUARES (IRLS)

Consider the linear model (KOCH 1999)

$$\mathbf{AX} = E(\mathbf{L}) \quad (1)$$

where  $\mathbf{A}$  is the  $n \times u$  design matrix with full column rank,  $\mathbf{X}$  the  $u \times 1$  vector of unknowns, and  $\mathbf{L}$  the  $n \times 1$  vector of independent and normally distributed observations with null mean vector and covariance matrix  $\sigma_0^2 \mathbf{P}^{-1} = \sigma_0^2 \text{diag}(1/p_i)$ .

The standard method for solving Eq. (1) is to compute a LS solution. However, the LS solution is very prone to outliers and even a single outlier will affect results considerably. One way to avoid this problem is to adopt the IRLS procedure, in which the discrepant observations are downweighted, rather than

merely deleted. Therefore, the IRLS algorithm is robust and the efficiency can be retained (HUBER 1981, CHANG and GUO 2005, HUBER and RONCHETTI 2009).

By choosing the a-priori weight matrix  $\mathbf{P}$  as the initial weight matrix  $\mathbf{P}^{(0)}$ , the IRLS scheme can be performed. Denoting the updated weight matrix with the  $k$ -th iteration by  $\mathbf{P}^{(k)} = \text{diag}(p_i^{(k)})$ , the estimate of unknowns is given by

$$\mathbf{X}^{(k)} = \mathbf{S}^{(k)} \mathbf{A}^T \mathbf{P}^{(k)} \mathbf{L} \quad (2)$$

with  $\mathbf{S}^{(k)} = (\mathbf{A}^T \mathbf{P}^{(k)} \mathbf{A})^{-1}$ .

The corresponding residual vector is readily obtained as

$$\mathbf{V}^{(k)} = \mathbf{L} - \mathbf{A} \mathbf{X}^{(k)} = \mathbf{R}^{(k)} \mathbf{L} \quad (3)$$

where  $\mathbf{R}^{(k)} = \mathbf{I} - \mathbf{A} \mathbf{S}^{(k)} \mathbf{A}^T \mathbf{P}^{(k)}$  is the reliability matrix or residual matrix (SCHAFFRIN 1997, GUO et al. 2007, 2011). It can be seen that  $\mathbf{R}^{(k)}$  is idempotent and thus, the sum of its diagonal elements  $r_{ii}^{(k)}$  is equal to the degree of freedom  $n - u$ . For uncorrelated observations,  $r_{ii}^{(k)}$  is called redundancy number and it holds that (SCHAFFRIN 1997, LEICK 2004)

$$0 \leq r_{ii}^{(k)} \leq 1 \quad (i = 1, 2, \dots, n) \quad (4)$$

### 3. VARIATION OF PRECISION AND DILUTION OF PRECISION METRICS

Without loss of generality, suppose the  $i_1$ -th, the  $i_2$ -th, ..., and the  $i_m$ -th observations be the  $m$  ( $< n - u - 1$ ) observations with reduced weights at the  $k$ -th step of the iteration. Let  $\mathbf{D}_m$  be an  $m \times m$  diagonal matrix whose diagonal entries are given by

$$d_j = p_{i_j}^{(k-1)} - p_{i_j}^{(k)}, \quad j = 1, 2, \dots, m \quad (5)$$

then we have

$$\mathbf{P}^{(k)} = \mathbf{P}^{(k-1)} - \mathbf{H}_m \mathbf{D}_m \mathbf{H}_m^T \quad (6)$$

with

$$\mathbf{H}_m = (\mathbf{h}_{i_1}, \mathbf{h}_{i_2}, \dots, \mathbf{h}_{i_m})$$

where  $\mathbf{h}_i$  represents the  $i$ -th  $n \times 1$  canonical unit vector with 1 in the  $i$ -th entry and zeros elsewhere. With the definition, it can be seen that the diagonal matrix  $\mathbf{D}_m$  is positive-definite.

With Sherman-Morrison-Woodbury-Schur formula (STRANG and BORRE 1997) and denoting  $\mathbf{A}_m = \mathbf{H}_m^T \mathbf{A}$ , we have

$$\begin{aligned} \mathbf{S}^{(k)} &= (\mathbf{A}^T \mathbf{P}^{(k-1)} \mathbf{A} - \mathbf{A}_m^T \mathbf{D}_m \mathbf{A}_m)^{-1} \\ &= \mathbf{S}^{(k-1)} + \mathbf{S}^{(k-1)} \mathbf{A}_m^T (\mathbf{D}_m^{-1} - \mathbf{A}_m \mathbf{S}^{(k-1)} \mathbf{A}_m^T)^{-1} \mathbf{A}_m \mathbf{S}^{(k-1)} \\ &= \mathbf{S}^{(k-1)} + \mathbf{S}^{(k-1)} \mathbf{A}_m^T \mathbf{B}_m \mathbf{A}_m \mathbf{S}^{(k-1)} \end{aligned} \quad (7)$$

where

$$\begin{aligned} \mathbf{B}_m &= (\mathbf{D}_m^{-1} - \mathbf{A}_m \mathbf{S}^{(k-1)} \mathbf{A}_m^T)^{-1} \\ &= \mathbf{D}_m + \mathbf{D}_m \mathbf{A}_m (\mathbf{A}^T \mathbf{P}^{(k-1)} \mathbf{A} - \mathbf{A}_m^T \mathbf{D}_m \mathbf{A}_m)^{-1} \mathbf{A}_m^T \mathbf{D}_m \\ &= \mathbf{D}_m + \mathbf{D}_m \mathbf{A}_m \mathbf{S}^{(k)} \mathbf{A}_m^T \mathbf{D}_m \end{aligned} \quad (8)$$

Obviously,  $\mathbf{B}_m$  is symmetric and positive-definite, since both  $\mathbf{D}_m$  and  $\mathbf{S}^{(k)}$  are symmetric and positive-definite.

The expression Eq. (7) shows the apparent decrease in precision when the downweighting procedure is performed. Moreover, all kinds of dilution of precision (DOP) metrics (STRANG and BORRE 1997) will become worse, but more realistic since  $[\mathbf{S}^{(k)}]_{ii} \geq [\mathbf{S}^{(k-1)}]_{ii}$  ( $i = 1, 2, \dots, n$ ).

#### 4. VARIATION OF MDB MEASURES

In reliability theory, MDB measures are used to describe the size of model errors that can be detected by simply using the appropriate test statistics (BAARDA 1968, TEUNISSEN 1985, VERHAGEN 2002, LEICK 2004). With Eq. (8), it can be verified that

$$\mathbf{B}_m \cdot (\mathbf{D}_m^{-1} - \mathbf{A}_m \mathbf{S}^{(k-1)} \mathbf{A}_m^T) \cdot \mathbf{D}_m = \mathbf{D}_m \quad (9)$$

which multiplies out to give

$$\mathbf{B}_m = \mathbf{D}_m + \mathbf{B}_m \mathbf{A}_m \mathbf{S}^{(k-1)} \mathbf{A}_m^T \mathbf{D}_m \quad (10)$$

and

$$\mathbf{B}_m = \mathbf{D}_m + \mathbf{D}_m \mathbf{A}_m \mathbf{S}^{(k-1)} \mathbf{A}_m^T \mathbf{B}_m. \quad (11)$$

By virtue of Eqs. (7) and (10), we have

$$\begin{aligned}
 AS^{(k)} A^T \cdot H_m D_m H_m^T &= A[S^{(k-1)} + S^{(k-1)} A_m^T B_m A_m S^{(k-1)}] A_m^T D_m H_m^T \\
 &= AS^{(k-1)} A_m^T (D_m + B_m A_m S^{(k-1)} A_m^T D_m) H_m^T \quad (12) \\
 &= AS^{(k-1)} A_m^T B_m H_m^T
 \end{aligned}$$

and

$$\begin{aligned}
 R^{(k)} &= I - AS^{(k)} A^T \cdot P^{(k-1)} + AS^{(k)} A^T \cdot H_m D_m H_m^T \\
 &= R^{(k-1)} - AS^{(k-1)} A_m^T B_m A_m S^{(k-1)} A^T P^{(k-1)} + AS^{(k-1)} A_m^T B_m H_m^T \quad (13) \\
 &= R^{(k-1)} + AS^{(k-1)} A_m^T B_m H_m^T \cdot R^{(k-1)}
 \end{aligned}$$

Taking Eq. (11) into account, one can obtain

$$\begin{aligned}
 H_m D_m H_m^T \cdot R^{(k)} &= H_m D_m H_m^T \cdot R^{(k-1)} + H_m D_m H_m^T \cdot AS^{(k-1)} A_m^T B_m H_m^T R^{(k-1)} \\
 &= H_m (D_m + D_m A_m S^{(k-1)} A_m^T B_m) H_m^T R^{(k-1)} \quad (14) \\
 &= H_m B_m H_m^T R^{(k-1)}
 \end{aligned}$$

which, in combination with Eq. (13) yields

$$\begin{aligned}
 P^{(k)} R^{(k)} &= (P^{(k-1)} - H_m D_m H_m^T) \cdot R^{(k)} \\
 &= P^{(k-1)} \cdot R^{(k-1)} + P^{(k-1)} \cdot AS^{(k-1)} A_m^T B_m H_m^T R^{(k-1)} - H_m B_m H_m^T R^{(k-1)} \quad (15) \\
 &= P^{(k-1)} R^{(k-1)} - (R^{(k-1)})^T H_m B_m H_m^T R^{(k-1)}
 \end{aligned}$$

This exact closed-form expression gives the relationship between  $P^{(k)} R^{(k)}$  and  $P^{(k-1)} R^{(k-1)}$ .

With noncentrality parameter  $\lambda_0$  (ibid), it follows from Eq. (15) that

$$\sigma_0 \sqrt{\frac{\lambda_0}{(P^{(k-1)} R^{(k-1)})_{ii}}} \leq \sigma_0 \sqrt{\frac{\lambda_0}{(P^{(k)} R^{(k)})_{ii}}}, \quad (i = 1, 2, \dots, n, k = 1, 2, \dots) \quad (16)$$

It can be seen that, all the MDB measures become larger and larger after performing the iterative reweighting procedure. These results coincide with those obtained using the first-order approximation method (GUO et al. 2011).

## 5. VARIATION OF BNR MEASURES

The external reliability expresses the effect of an undetected error on the final estimation results. In practical applications, external reliability is usually much more

relevant than internal reliability (BAARDA 1968, TEUNISSEN 1985, VERHAGEN 2002, LEICK 2004).

The bias-to-noise ratio (BNR) measure is one of two scalar external reliability measures. Under the diagonality assumption of the weight matrix, the  $i$ -th BNR measure with the  $k$ -th iteration is defined as (BAARDA 1968, TEUNISSEN 1985, VERHAGEN 2002, LEICK 2004)

$$\text{BNR}_i = \sqrt{\lambda_0 (1/r_{ii}^{(k)} - 1)} \tag{17}$$

Suppose the relationship between the weight elements  $p_i^{(k)}$  and  $p_i^{(k-1)}$  is as follows

$$p_i^{(k)} = f_i p_i^{(k-1)} \tag{18}$$

in which the factor  $f_i$  ( $0 < f_i \leq 1$ ) is the function of the  $i$ -th (standardized) residual obtained in the  $(k - 1)$ -th-iteration. Determination of  $f_i$  is termed as downweighting strategy, which has attracted a great deal of attention both in statistical and geodetic literature (HUBER 1981, CHANG and GUO 2005, HUBER and RONCHETTI 2009, GUO et al. 2010, 2011).

With Eq. (16), if  $p_i^{(k)} = p_i^{(k-1)}$ , then

$$r_{ii}^{(k)} \leq r_{ii}^{(k-1)} \tag{19}$$

For the subsequent discussions we introduce the following theorem:

**Theorem:** Assume that both  $\mathbf{M}$  and  $\mathbf{N}$  are  $n \times n$  symmetric positive definite matrices, then  $\mathbf{M} - \mathbf{N}$  is positive semi-definite if and only if  $\mathbf{N}^{-1} - \mathbf{M}^{-1}$  is positive semi-definite.

For the proof of this theorem, the reader is referred to Horn and Johnson (1985, p 471).

In case of  $f_j = \min_i f_i$ , then  $[f_j \mathbf{S}^{(k)}]^{-1} - [\mathbf{S}^{(k-1)}]^{-1} = \mathbf{A}^T \text{diag}((f_j - f_j) p_i^{(k-1)}) \mathbf{A} / f_j$  is positive semi-definite. According to the above theorem, one can conclude that  $\mathbf{S}^{(k-1)} - f_j \mathbf{S}^{(k)}$  is also positive semi-definite and thus

$$\begin{aligned} r_{jj}^{(k)} - r_{jj}^{(k-1)} &= [1 - \mathbf{h}_j^T \mathbf{A} \mathbf{S}^{(k)} \mathbf{A}^T \mathbf{P}^{(k)} \mathbf{h}_j] - [1 - \mathbf{h}_j^T \mathbf{A} \mathbf{S}^{(k-1)} \mathbf{A}^T \mathbf{P}^{(k-1)} \mathbf{h}_j] \\ &= \mathbf{h}_j^T \mathbf{A} (\mathbf{S}^{(k-1)} - f_j \mathbf{S}^{(k)}) \mathbf{A}^T \mathbf{h}_j \cdot p_j^{(k-1)} \geq 0 \end{aligned} \tag{20}$$

However, if  $\min_i f_i < f_j < 1$ , the sign of the expression  $r_{jj}^{(k)} - r_{jj}^{(k-1)}$  is ambiguous.

Therefore, at any two consecutive iteration steps, (1) the BNR measures of observations whose weights keep unchanged become larger; (2) the BNR measure of the observation with maximum absolute standardized residual decreases, whereas the BNR measures of other observations with reduced weights may become larger or smaller. These results also coincide with those obtained by the first order approximation method (GUO et al. 2011).

**6. A NUMERICAL EXAMPLE**

A simulated geodetic leveling network as shown in Figure 1 was taken as a test example. The elevation of station A is 168.0000 m. The simulated observations and their weights are listed in Table 1.

Figure 1 - Leveling network.

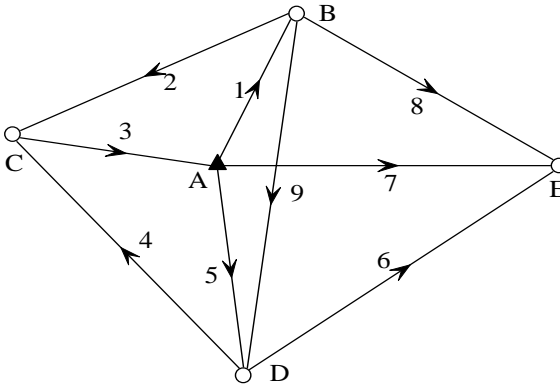


Table 1 -. Simulated leveling network.

Line	Leveling difference (m)	Relative weight
1	5.7886	1
2	-14.4582	2
3	8.6681	2
4	-12.3387	1
5	3.6702	2
6	-14.3008	1
7	10.6318	2
8	-16.4200	1
9	-2.1178	2

For purpose of illustration, two artificial outliers,  $-0.08$  and  $+0.06$  (m) are added to the third and the ninth observation, respectively. The damping factors can be determined as follows (GUO et al. 2011)

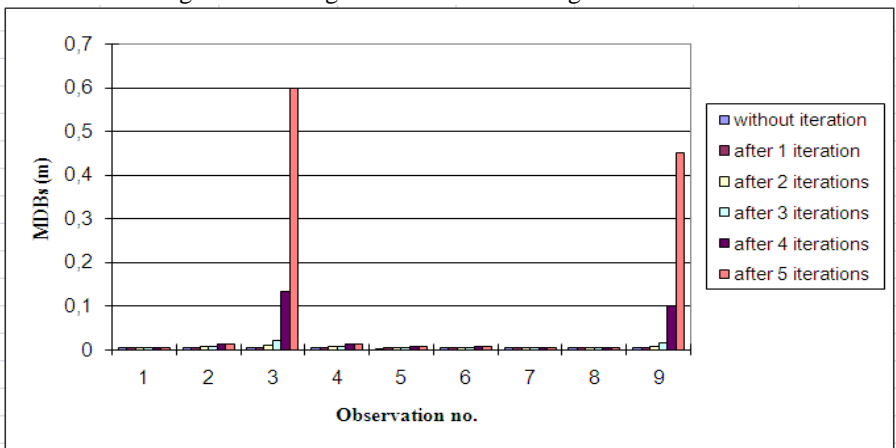
$$f_i = \begin{cases} 1, & |\tilde{v}_i| \leq c \\ (c/|\tilde{v}_i|)^2, & |\tilde{v}_i| > c \end{cases}, (i = 1, 2, \dots, n) \quad (21)$$

where  $\tilde{v}_i$  is the  $i$ th standardized LS residual and the constant  $c$  is usually taken from the interval  $[1.5, 2.0]$  (KOCH 1999, CHANG and GUO 2005). If the difference between the estimated unknowns at two consecutive iterations is less than a positive constant, or the number of iterations surpasses a preset threshold number, then the iteration process should be stopped.

The median absolute deviation (MAD) estimate is the candidate for being the “most robust estimate of scale” (HUBER 1981). However, in order to make MAD consistent at the normal distribution, we must multiply it by 1.4826. Therefore, the scale factor involved in  $\tilde{v}_i$  can be replaced by its normalized MAD estimate (ROUSSEEUW and LEROY 1987; GUO et al. 2010, 2011).

In geodetic applications (BAARDA 1968), the significance level and the detection power are commonly set at 0.001 and 0.80, respectively. This results in a noncentrality parameter  $\sqrt{\lambda_0} = 4.13$ . The parameter  $\sigma_0$  used in the MDB and BNR measures is taken as 0.001 m. The aforementioned convergence criterion meets after five iterations. As expected, all the MDB measures become larger and larger during the iteration procedure (cf. Figure 2).

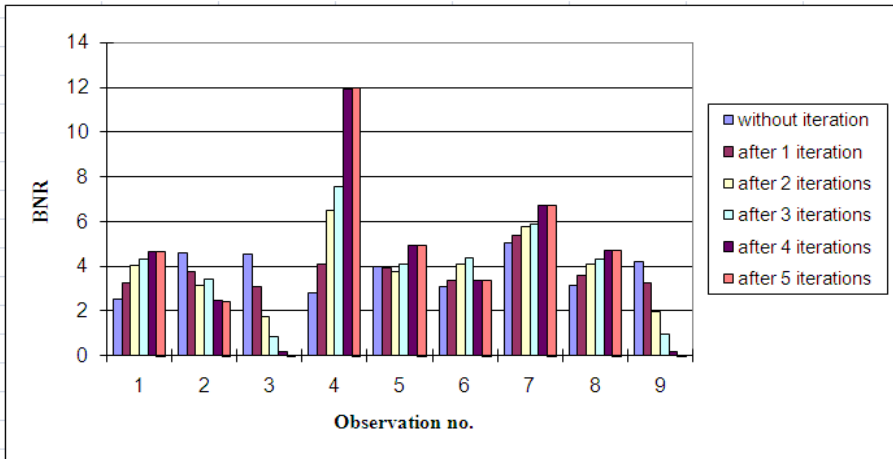
Figure 2 - Changes in the MDBs during the iterations.





When the stopping criterion meets, the BNR measures corresponding to the two outlying observations are considerably small (see Figure 3).

Figure 3 - Variation of the BNRs over the five iterations.



These results can be used to explain how the IRLS technique resists the presence of outliers and mitigate their impact on the final estimated parameter.

## 7. CONCLUSIONS

The IRLS technique has been extensively employed in geodetic and geophysical literature. To gain insight into the IRLS method, an exact and direct analytical method is presented to obtain insights on how much the iterative reweighting can impair the quality indicators. Theoretical analyses and numerical results show that, when the downweighting procedure is performed, (1) the precision, all kinds of DOP metrics and MDB measures will become larger; (2) the variations of BNR measures are ambiguous, and (3) all these results coincide with those obtained by the first-order approximation method (GUO et al. 2011).

## ACKNOWLEDGEMENTS

This research was sponsored by National Key Basic Research Program of China (2012CB825604), and the Natural Science Foundation of China (Grant Nos. 41374041 and 40874007). The author is also supported by the China Scholarship Council (File No. 2011317045).

## REFERENCES

BAARDA W. A testing procedure for use in geodetic networks. *Netherlands Geod. Comm., Publ. on Geodesy*, 1968, New Series, 2(5), Delft, The Netherlands.

- CHANG X.; GUO Y. Huber's M-estimation in relative GPS positioning: computational aspects. *Journal of Geodesy*, 79(6-7): 351-362, 2005.
- COLLILIEUX X.; VAN DAM T.; RAY J.; COULOT D.; MÉTIVIER L.; ALTAMIMI Z. Strategies to mitigate aliasing of loading signals while estimating GPS frame parameters. *Journal of Geodesy*, 86(1): 1-14, 2012.
- GUO J. The case-deletion and mean-shift outlier models: equivalence and beyond. *Acta Geod. Geophys.*, 48(2): 191-197, 2013.
- GUO J.; OU J.; WANG H. Quasi-accurate detection of outliers for correlated observations. *Journal of Surveying Engineering*, 133(3): 129-133, 2007.
- GUO J.; OU J.; WANG H. Robust estimation for correlated observations: two local sensitivity-based downweighting strategies. *Journal of Geodesy*, 84(4): 243-250, 2010.
- GUO J.; OU J.; YUAN Y. Reliability analysis for a robust M-estimator. *Journal of Surveying Engineering*, 137(1): 9-13, 2011.
- HORN R. A.; JOHNSON C. R. Matrix Analysis. Cambridge Univ. Press, Cambridge, UK, 1985.
- HUBER P. J. Robust statistics, Wiley, New York, 1981.
- HUBER P. J.; RONCHETTI E. M. Robust statistics, 2nd Ed., Wiley, New York, 2009.
- KNIGHT N. L.; WANG J.; RIZOS C. Generalised measures of reliability for multiple outliers. *Journal of Geodesy*, 84(10): 625-635, 2010.
- KOCH K. R. Parameter estimation and hypothesis testing in linear models, 2nd Ed., Springer, Berlin, 1999.
- LEICK A. GPS satellite surveying, 3rd Ed., Wiley, New York, 2004.
- OU J. On the reliability for the situation of correlated observations. *Acta Geodaet. et Cartograph. Sinica, English Edition*, 9-17, 1999.
- RANGELOVA E.; FOTOPoulos G.; SIDERIS M. G. On the use of iterative re-weighting least-squares outlier detection for empirically modelling rates of vertical displacement. *Journal of Geodesy*, 83(6): 523-535, 2009.
- ROUSSEUW P. J.; LEROY A. M. Robust regression and outlier detection. Wiley, New York, 1987.
- SCHAFFRIN B. Reliability measures for correlated observations. *Journal of Surveying Engineering*, 123(3): 126-137, 1997.
- STRANG G.; BORRE K. Linear algebra, geodesy, and GPS. Wellesley-Cambridge Press, Wellesley, 1997.
- TEUNISSEN P. J. G. Quality control in geodetic networks. In: Optimization and design of geodetic networks, E. W. Grafarend and F. Sanso, eds., Springer, Berlin, 526-547, 1985.
- VERHAGEN S. Studying the performance of Global Navigation Satellite Systems: A new software tool. *GPS World*, 13(6): 60-65, 2002.
- WANG J.; CHEN Y. On the reliability measure of observations. *Acta Geodaet. et Cartograph. Sinica, English Edition*, 42-51, 1994.
- (Recebido em julho de 2013. Aceito em novembro de 2013).