e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:02/February-2021 Impact Factor- 5.354 www.irjmets.com

EFFECT OF REDUNDANCY ON GROSS ERROR DETECTION

Khalid Ali Mohammed Haidar*1, Ahmed Mohmed Ibrahim*2

*1Dept. of Civil Engineering, Faculty of Engineering Science, Nyala University, Sudan.

*2Dept. of Surveying Engineering, College of Engineering, Sudan University of Science and Technology, Sudan.

ABSTRACT

Most of the surveying tasks involve acquisition and analysis of measurements. The method of least squares estimation is commonly used to process the measurements. In practice, redundant measurements are made to provide quality control and check the presence of errors that might affect the results. Therefore, an insurance of the quality of these measurements is an important issue. Measurement errors of collected data have different levels of influence due to their number, measured accuracy and redundancy. This paper intends to explore the issue of gross error detection in horizontal control networks. The methods of detecting these gross errors are used in conjunction with developed programs to calculate critical values for the distributions (in real time) rather than looking for them in statistical tables. The main conclusion reached is that the tau (τ) statistic is the most sensitive in detecting the presence of gross errors with the least number of redundant observations; therefore, it is the one recommended to be used in gross error detection in horizontal control networks.

Key words: gross error, statistical test, data snooping, redundancy, quality control.

I. INTRODUCTION

Gross errors are the result of a malfunctioning of either the instrument or the surveyor. Typical examples are the incorrect reading or incorrect recording of results and failure of the instrument due to weak power supply or extreme environmental conditions. At least, theoretically, gross errors can be avoided by due care or they can be detected by carefully designed observation schemes. (Caspary, 1987).

For high precision applications such as deformation monitoring, gross errors need to be detected and localized prior to deformation analysis. Whenever possible, gross errors should be tackled before **L**east **S**quares **E**stimation (**LSE**), by means of screening and independent checks (Cooper, 1974, 1987; Secord, 1986; Ibrahim, 1995).

This paper aims to explore errors detection capability of three schemes; namely, the Global test, data snooping and Tau test, in order to compare the characteristics of the three methods of gross error detection. With the least squares method, when there are gross errors in the observations, the magnitudes of corresponding residuals may not always be larger than other residuals having no gross errors. This makes gross errors difficult to be found (detected). Therefore, it is not reliable if gross errors are to be detected by simply examining the magnitudes of the residuals alone.

II. LEAST SQUARES ESTIMATION (LSE)

The Least Squares method is only one of the estimation techniques used in surveying. It has been widely used in most practical situations because of its simplicity and also because complete statistical information is generally available as a result. In addition, it gives estimated values which are statistically equal to their true values (unbiasedness). Also, it gives variances which are smaller than the variances resulting from any other estimation method (efficiency). For these last two reasons, LSE is considered to be the most used method of estimation.

The main equations for LSE using observation equations method are shown here without further derivation or proof. More details are found extensively in surveying literature for example (Cross, 1983). The fundamental Equations for LSE with n observations, m parameters and redundancy r are as follows:

$$\hat{x} = (A^T W A)^{-1} A^T W b \quad \text{............ (1)}$$

$$\hat{v} = A\hat{x} - b \quad \text{............ (2)}$$

$$\hat{\sigma}_0^2 = \hat{v}^T W \hat{v} / r \quad \text{............ (3)}$$

$$C_{\hat{v}} = W^{-1} - C_{\hat{l}} \quad \text{.......... (4)}$$

$$C_{\hat{l}} = A(A^T W A)^{-1} A^T \quad \text{.............. (5)}$$

$$r = n - m \quad \text{............ (6)}$$

Where:

$\hat{v}$: Vector of estimated residuals

W: weight matrix

A: the design (coefficient) matrix

$\hat{x}$: estimated parameters

$C_{\hat{v}}$: Covariance matrix of the residuals

$C_{\hat{l}}$: Coefficient matrix of the observations after LSE.

b: vector of the difference between observed values and corresponding computed values using approximate values for the parameters.

III. GROSS ERROR DETECTION METHODS

In recent years, the detection of gross errors and the reliability of observations have been one of the main research directions in surveying.

We need to have some method of analyzing the results of a least squares computation to determine whether or not any of the observations are outliers. These methods depend on the analysis of residuals after an estimation process has been carried out. If we assume that the observed quantities are normally distributed which are generally so, then the residuals of these observations are also normally distributed with zero mean because the least squares method tends to minimize the weighted sum of residuals.

The gross error in an observation usually affects the residuals in other observations. Therefore, if an observation does not pass a statistical test, this does not mean that there is a gross error in that observation. Thus, a statistical test should be applied to detect large errors or mistakes.

Gross error detection methods applied in this paper comprise Global test (G), data snooping (ω) and Tau (τ) tests.

3.1 The Global Test (G)

The first test which is applied after any estimation process is the well known global test on the a posteriori variance factor $\hat{\sigma}_0^2$. This test can obviously be applied only when there is a priori knowledge about the precision of the observations, i.e. when the a priori variance factor σ_0^2 is assumed to be known. Otherwise the test has no meaning.

Under the null hypothesis H_0 the statistic $\hat{\sigma}_0^2/\sigma_0^2$ follows the F – distribution with r and ∞ degree of freedom. It is to be remembered that $F_{r,\infty} = \chi_r^2/r$. The decision for this global test (one-tailed or two-tailed) depends on the purpose of the test which is defined by the null hypothesis H_0.

The two- tailed test takes the form:

$$H_0: \hat{\sigma}_0^2 = \sigma_0^2$$

$$H_a: \hat{\sigma}_0^2 \neq \sigma_0^2$$

Where, σ_0^2 represents the variance factor, or the variance of unit weight as is sometimes called, and $\hat{\sigma}_0^2$ is the estimated value of the variance factor. This gives the following $100(1-\alpha)\%$ confidence interval for the variance factor σ_0^2 :

$$P\left[\frac{r\hat{\sigma}_0^2}{\chi^2_{r,1-\alpha/2}} \leq \sigma_0^2 \leq \frac{r\hat{\sigma}_0^2}{\chi^2_{r,\alpha/2}}\right] = 1-\alpha \quad \text{...... (7)}$$

When the global test is used for the detection of gross errors, it is normally expected that $\hat{\sigma}_0^2$ will be greater than σ_0^2. Therefore, a one-tailed right hand test is recommended which takes the form:

$$H_0: \hat{\sigma}_0^2 \geq \sigma_0^2$$
$$H_a: \hat{\sigma}_0^2 < \sigma_0^2$$

and the test statistic follows an F distribution: i.e

$$\frac{\hat{\sigma}_0^2}{\sigma_0^2} \sim F_{1-\alpha;r,\infty} \quad \text{............ (8)}$$

Since $\hat{\sigma}_0^2/\sigma_0^2 = \hat{v}^T W\hat{v}/r\sigma_0^2$, equation (8) can be written as:

$$\hat{v}^T W\hat{v}/r\sigma_0^2 \sim F_{1-\alpha;r,\infty} \quad \text{............ (9)}$$

Since $\frac{\chi^2_{1-\alpha,r}}{r} = F_{1-\alpha,r,\infty}$, then

$$\hat{v}^T W \hat{v} / \sigma_0^2 \sim \chi^2_{1-\alpha;r} \quad \text{............ (10)}$$

The use of either equation (9 or 10) depends on the researcher.

If there are any gross errors in the data then the above quadratic form will increase and the test may fail or not, depending on the magnitudes of the gross errors and on how they are reflected in the residuals. If this test (equation (8)) fails, then H_0 is rejected. Unfortunately, there may be more than one reason for rejection (Uotila , 1976); for example:

1. An incorrect estimation of weights;
2. Incorrect mathematical model;
3. There are gross errors in the observations.

We may not know which one of the above reasons caused the failure of the test, and the test does not give any additional information. Whatever the reason is, it should be investigated, thoroughly, and not simply ignored. Confining ourselves to the third possible cause for rejection, i.e. gross errors in the observations, an alternative hypothesis, H_a can be introduced, see (Van Mierlo , 1977)

3.2 Data Snooping (ω - Test)

The theory of this technique is developed and introduced by professor Baarda of the Netherland (Baarda, 1968) for use in geodetic control networks. Assuming that the residuals indicate linear functions of observations, the normalized residuals are used for evaluation.

This method first utilizes a test of a global model, e.g., using the statistic $\hat{\sigma}_0^2 / \sigma_0^2$ as described in section (3.1). When this statistic is less than the threshold, the global model is considered correct, i.e. no major errors exist in the observations. In other words, mistakes do not exist in the observations. The threshold value is obtained from F - distribution with r and ∞ degree of freedom with the commonly applied level of significance α .

Baarda proposed the use of the global test (9) for the detection of gross errors and the "Data Snooping" test (11) for their localization. The decisions from both tests should be consistent, i.e., the same boundary values should be found whether the global or the single, ω, test is performed.

For detecting each individual observation, the residuals and α can be standardized to obtain a standardized residual ω_i and standardized α_0 as follows:

$$\omega_i = \left| \frac{\hat{v}_i}{\sigma_{\hat{v}_i}} \right| \sim \sqrt{F_{1-\alpha_0;1,\infty}} \quad \text{......... (11)}a$$

Which follows a standardized normal distribution ($N(0,1)$) i.e

$$\omega_i = \left| \frac{\hat{v}_i}{\sigma_{\hat{v}_i}} \right| \sim N_{1-\alpha_0/2} \quad \text{......... (11)}b$$

$$\alpha_0 = 1-(1-\alpha)^{1/n} \cong \alpha / n \quad \text{.......... (12)}$$

IRJMETS

e-ISSN: 2582-5208

International Research Journal of Modernization in Engineering Technology and Science

Volume:03/Issue:02/February-2021 Impact Factor- 5.354 www.irjmets.com

Where: $\sigma_{\hat{v}_i}$ is the aposterior standard deviation of the residual given by the square root of the *ith* diagonal element of matrix $C_{\hat{v}}$ in (4).

The test can be applied as follows:

1. The least squares estimation is used to estimate $\hat{v}$ and $C_{\hat{v}}$ from (2) and (4) respectively.
2. The level of significance α is determined and standardized to α_0 using (12).
3. The critical value ω_c is determined from the available program written for this purpose using the level of significance α_0.
4. The statistic ω_i is computed for each observation using (11)a.
5. The computed value, ω_i, is compared with the critical value, ω_c
6. Check if the maximum standardized residual does not reflect the presence of any gross error. i.e if $\omega_i \le \omega_C$.

Otherwise remove the observation containing a gross error and repeat until all data is screened.

Baarda's method, assumes that σ_0^2 is known apriori, and employs a multi dimensional test. In the actual implementation of Baarda's method, both *Type*I and *Type*II errors should be taken into account.

3.3 The Tau Test (τ)

In the null hypothesis of the previous tests the variance of unit weight, σ_0^2 is assumed to be known. This means that all variances are properly scaled. If, however, σ_0^2 is not adequately known or one does not want to rely on a priori estimates, then the a posteriori estimate $\hat{\sigma}_0^2$ is always available from LSE. In this case, the global test on the variance is not performed and the method of Data Snooping has to be modified. The new test statistic, proposed by Pope (Pope, 1976), which takes the form given bellow, is the one to be used.

$$\tau_i = \frac{\hat{v}_i}{\hat{\sigma}_0 \sigma_{\hat{v}_i}} = \frac{\omega_i}{\hat{\sigma}_0} \qquad \text{............ (13)}$$

This statistic follows the so called Tau distribution. Since the residuals are used for the estimation of τ statistic through $\hat{\sigma}_0^2$ Pope's, or Tau, method assumes σ_0^2 as unknown and applies its least square estimate in computing the normalized residuals. The test statistic is one dimensional i.e

$$\tau_i = \left| \frac{\hat{v}_i}{\hat{\sigma}_0 \sigma_{\hat{v}_i}} \right| \sim \tau_{1-\alpha;n,r} \qquad \text{............ (14)}$$

where:

$$\alpha = n\alpha_0 \% \qquad \text{.......... (15)}$$

It should be noted that this test is a one-tailed, left hand, test. That is H_0 is accepted if:

$$\tau_i \le \tau_{1-\alpha;n,r} \qquad \text{............ (16)}$$

Otherwise H_0 is rejected, and the corresponding observation is suspected of having a gross error, provided the mathematical model is correct and the weights are properly determined. Unlike the ω test, this test does not take into account the probability of *Type*II error.

The Tau test can be set up using least squares results as follows:

1. The least squares estimation is used to estimate $\hat{v}$ and $C_{\hat{v}}$ from (2) and (4) respectively.
2. The level of significance α is determined and standardized to α_0 using (12).
3. The critical value τ_c , is determined from the developed program using the level of significance $n\alpha_0\%$.
4. The statistic τ_i is computed for each observation using (13).
5. The computed value, τ_i , is compared with the critical value, τ_c .
6. Check whether H_0 is accepted or not; if not then, that indicates the presence of a gross error in that observation, otherwise it is not.
7. Remove the observation having a gross error and repeat the test for the remaining observations until all data is screened.

VI. TESTS AND DISCUSSION

The following test illustrates and demonstrates the effect of redundancy on gross error detection. The following measurements and data pertain to the triangulation chain shown in Figure (1) below with coordinates of control stations, approximate coordinates of unknown stations, and measurements as shown in Tables (1), (2) and (3) respectively.

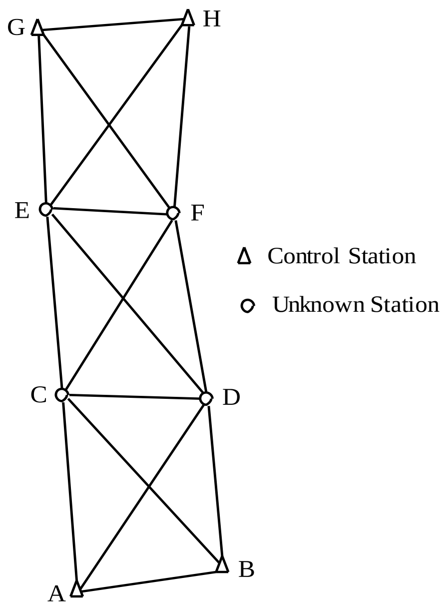

Figure 1: The Horizontal Network

Table 1: Control Stations Coordinates

Station	*E(m)*	*N(m)*
A	1718.871	632.095
B	2191.570	715.709

G	1590.370	2560.743
H	2076.006	2597.745

Table 2: Approximate Coordinates of Unknown Stations

Station	*E(m)*	*N(m)*
C	1668.571	1310.429
D	2139.109	1296.242
E	1617.479	1949.217
F	2028.688	1934.566

Table 3: Observed Angles

Back Sight	*Occupied*	*Foresight*	*Angle*	*σ*
C	A	D	36° 33′ 49.5″	4.6″
D	A	B	47° 38′ 40.5″	5.7″
A	B	C	58° 42′ 5.4″	5.7″
C	B	D	36° 9′ 56.9″	4.9″
B	C	A	37° 5′ 18.5″	4.6″
D	C	B	46° 56′ 40.5″	5.7″
B	D	A	37° 29′ 8.7″	5″
A	D	C	59° 24′ 14.1″	5.8″
E	C	F	34° 33′ 22.5″	4.8″
F	C	D	61° 44′ 34.1″	5.9″
C	D	E	49° 39′ 14.4″	5.7″
E	D	F	28° 48′ 19.6″	4.6″
D	E	C	49° 20′ 57.4″	4.6″
D	F	C	34° 02′ 48.2″	4.8″
C	F	E	39° 47′ 51.2″	6.5″
G	E	H	62° 03′ 22.6″	4.8″

IRJMETS
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:02/February-2021 Impact Factor- 5.354 www.irjmets.com

H	E	F	$37^{o}\ 47'\ 58.1''$	6.3”
E	F	G	$56^{o}\ 46'\ 43.5''$	6.4”
G	F	H	$52^{o}\ 58'\ 5.9''$	4.7”
F	G	E	$39^{o}\ 04'\ 18.8''$	4.8”
H	G	F	$32^{o}\ 27'\ 12.1''$	5.7”
F	H	E	$59^{o}\ 21'\ 51.3''$	4.6”
E	H	G	$31^{o}\ 10'\ 44.3''$	5.6”
mean = $\overline{\sigma}$ = 5.3”				

The effect of redundancy on the possibility of detecting gross errors is to be investigated using the three statistics mentioned earlier.

From these three tables, the weight matrix (W), coefficient matrix (A), and the vector of observed values minus their respective values computed from approximate coordinates (b) are determined and used for the LSE (Equations (1) through to (6)). The results of LSE are then used to calculate the three statistics, namely τ, ω, and *G* statistics.

At the beginning, all these statistics were calculated and determined using the approximate values obtained from the actual observations which have no gross errors, and the variance factor which was found to be equal to 0.004. Gross errors ranging from $2.7\overline{\sigma}$ to $45\overline{\sigma}$ seconds were then added to the first observation, and the three statistics calculated using a signeficance level of 5% (95% probability). It is found that the τ statistic can detect gross errors as small as $2.7\overline{\sigma}$. On contrast the global test can only detect gross errors of size $45\overline{\sigma}$ seconds (approximate $00^{o}\ 03'\ 58''$), or larger, when the redundancy is (15). The ω test is a little bit better than the global *G* statistic; it can only detect gross errors as large as $28\overline{\sigma}$ seconds (approximate $00^{o}\ 02'\ 28''$) when the redundancy is (15). Therefore, these two methods (the G and ω) are not investigated any further and the most sensitive method of the three is the one used.

Gross errors of size ranging from $2.7\overline{\sigma}$ to $8\overline{\sigma}$ seconds, in steps of $0.5\overline{\sigma}$ seconds were added, one after another, to the first observation. The calculated τ statistic for redundancies ranging from 12 down to 4 were determined and found to be as those in Table (4) and shown, graphically, on Figure (2).

Table 4: The Relation between Gross Errors, $\hat{\sigma}_0^2$ and Redundancies

Size of detectable gross error	*Minimum **r***	**τ** *Critical Value*	**τ** *Test Value*	$\hat{\sigma}_0^2$
2.7σ̄	12	2.633	2.644	0.008
3 σ̄	10	2.543	2.597	0.010
3.5 σ̄	9	2.488	2.584	0.014

4 σ̄	7	2.344	2.375	0.019
4.5 σ̄	6	2.245	2.261	0.028
5 σ̄	6	2.245	2.299	0.034
5.5 σ̄	5	2.120	2.124	0.050
6 σ̄	5	2.120	2.143	0.060
6.5 σ̄	5	2.120	2.158	0.071
7 σ̄	5	2.120	2.169	0.084
7.5 σ̄	5	2.120	2.178	0.096
8 σ̄	4	1.952	1.955	0.138

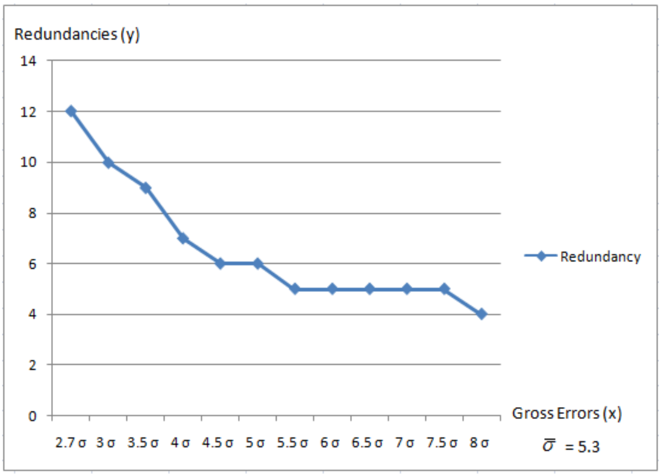

Figure 2: The Relation between Gross Errors and Redundancies

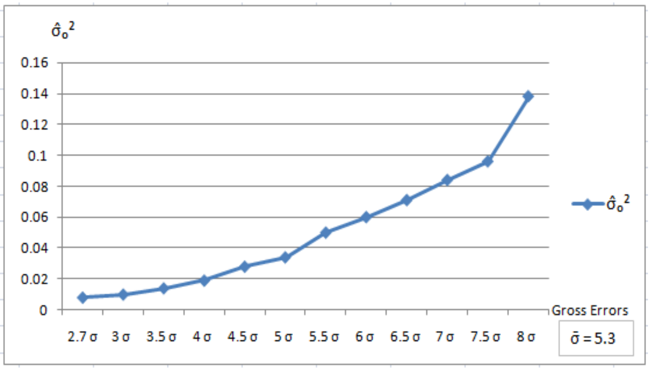

Figure 3: The Relation between Gross Errors and $\hat{\sigma}_0^2$

4.1 Discussion:

It can be seen from Table (4) and Figure (2) that in the presence of a gross error, the minimum redundancy to detect gross error decreases with an increase in the size of a gross error. It decreases from 12 to 4 when the size of a gross error increased from 2.7 $\overline{\sigma}$ to 8 $\overline{\sigma}$ seconds. This is to be expected since the larger is the size of a gross error then it is more reflectable in the test statistic.

As has been pointed out earlier this is an indication that the τ statistic is the most sensitive to gross errors compared to the other two statistics in terms of redundancy. Very small errors can be reflected in the τ statistic and can, therefore be detected.

Comparing the size of a gross error that any method can detect, with a probability of 0.95 (significance level of 0.05) is approximately 45 $\overline{\sigma}$, 28 $\overline{\sigma}$, and 2.7 $\overline{\sigma}$ seconds of arc for global test (G), Baarda's data snooping (ω), and Pope's method (τ) respectively. This result conforms to and agrees with the foregoing result discussed in the previous paragraph. Namely, the τ statistic is the most sensitive of the three statistics and is, therefore, the one recommended to be used in gross error detection. It can detect gross errors as small as 2.7 $\overline{\sigma}$ seconds of arc.

These resulting are used to derive an empirical model that relates the number of redundant observations and the size of a gross error that can be detected as given by equation (17) bellow:

$$y = 0.013x^2 - 1.007x + 22.83 \quad \text{..........} \quad (17)$$

where:

y: is the redundancy

x: size of gross error

Also, from Table (4), it can be seen that $\hat{\sigma}_0^2$ increase with an increase in the size of a gross error. It changes from 0.008 when the size of gross error is 2.7 $\overline{\sigma}$ seconds, to 0.138 when the size of gross error is increased to 8 $\overline{\sigma}$ seconds.

IV. CONCLUSIONS

From the tests carried out and the results obtained, the following conclusions could be drawn:

- The variance factor increases in size with an increase in the size of a gross error. Its size reflects whether there is a gross error or not. The variance of unit weights increases faster than the size of a gross error. When the size of a gross error is increased 4 times, the variance of unit weight increased approximately 17 times.
- The τ (tau) statistic is the most sensitive to gross errors compared to the other two statistics (G and ω). Errors as small as (2.7 $\overline{\sigma}$ seconds) can be detect using the τ statistic.
- The ω statistic is better than the global test statistic in gross error detection.

V. REFERENCES

[1] Achtert, E., Kriegel, H.-P., Reichert, L., Schubert, E., Wojdanowski, R., Zimek, A. (2010). "Visual Evaluation of Outlier Detection Models." In Proc. International Conference on Database Systems for Advanced Applications (DASFAA), Tsukuba, Japan.

[2] Barada, W., (1968), "A testing Procedure for use in Geodetic Networks" Publications on Geodesy, New Series, Netherlands Geodetic Commission, V. 2, N. 5, Delft, The Netherlands.

[3] Caspary, W.F 1987,"Concepts of networks and deformation analysis" School of surveying monograph II,

University of New South Wales, 183 PP.

[4] Cooper, M.A.R 1987 "Control Surveying In Civil Engineering" Blackwell Scientific Oxford, 381 PP.

[5] Cross P.A 1983 "Advanced Least Squares Applied To Position Fixing," Working Paper No6 department of surveying, North East London polytechnic, 185 PP.

[6] Ibrahim, A. M. (1995), Reliability Analysis of Combined GPS-Aerial Triangulation System, Ph.D. Thesis, Newcastle University, England.

[7] MIERLO, J. van (1977) Systematic Investigations on the Stability of Control Points, presented paper at XV International Congress of Surveyors, Stockholm, Sweden.

[8] Pope, A, 1976, "The Statistics of Residuals and the Detection of Outliers" NOAA Technical Report NOS 65 NGS 1, U.S. Department of Commerce, Washington, D.C., USA.

[9] Secord. J. M. (1986) "Terrestrial Survey Methods for Preston Deformation Measurements" Oct. 31-Nov. I. Massachusetts Institute of Technology, Cambridge, MA

[10] Shen Y., Li B. and Chen Y., 2011, An Iterative Solution of Weighted Total Least-Squares Adjustment, J Geod, 85,4, 229-238.

[11] Tong X., Jin Y. and Li L., 2011, An Improved Weighted Total Least Squares Method with Applications in Linear Fitting and Coordinate Transformation, J Surv. Eng., 137,4, 120-128.

[12] UOTILA, A.U. (1976) Statistical Tests as Guidelines in Analysis of Adjustment of Control Nets, presented paper, Federation International des Geometries, 14th Congress, Washington, D.C.

[13] Wang J, Wang J (2007) Mitigation the Effects of Multiple Outliers on GNSS Navigation with M-Estimation Schemes. In: IGNSS Symposium 2007. 4-6 December 2007, Sydney, 1-9.

[14] Yang, L., Wang, J., Knight, N.L. and Shen, Y., 2013. Outlier Separability Analysis with a Multiple Alternative Hypotheses Test. Journal of Geodesy, 87(6), pp. 591–604.