

# UČINKOVITOST RAZLIČNIH NAČINOV ODKRIVANJA GROBIH POGREŠKOV V GEODETSKIH MREŽAH: PRIMER SPOMENIKA POBEDNIK

# THE EFFICIENCY OF DIFFERENT OUTLIER DETECTION APPROACHES IN GEODETIC NETWORKS: CASE STUDY FOR POBEDNIK STATUE

Mehmed Batilović, Zoran Sušić, Marko Z. Marković, Marijana Vujinović, Gojko Nikolić, Toša Ninkov

UDK: 528.1:528.41(497.11 Beograd)  
Klasifikacija prispevka po COBISS.SI: 1.01  
Prispelo: 12. 12. 2017  
Sprejeto: 25. 5. 2018

DOI: 10.15292/geodetski-vestnik.2018.02.293-305  
SCIENTIFIC ARTICLE  
Received: 12. 12. 2017  
Accepted: 25. 5. 2018

## IZVLEČEK

V članku obravnavamo učinkovitost odkrivanja grobih pogreškov opazovanj v geodetskih mrežah za dve različni metodi. Analiza je izvedena na primeru kontrolne mreže za geodetska opazovanja spomenika Pobednik na območju trdnjave Kalemegdan v Beogradu, kjer je uporabljena ocena povprečne uspešnosti (angl. MSR – mean success rate). Za odkrivanje grobih pogreškov smo uporabili klasične analize in robustne metode. Rezultati empirične analize originalnih geodetskih opazovanj so pokazali, da so robustne metode učinkovitejše od klasičnih analiz odkrivanja grobih napak opazovanj v geodetskih mrežah.

## ABSTRACT

In the paper the efficacy of two different approaches for outlier detection in geodetic networks is analysed on a test example of a control network for geodetic monitoring of the Pobednik statue in the Kalemegdan Fortress in Belgrade by applying the mean success rate (MSR). Conventional tests and robust methods were applied for detecting outliers. The experimental results indicate that the new approach based on original observations provides higher efficiency of the applied methods than the classical approach for outlier detection in geodetic networks.

## KLJUČNE BESEDE

grobi pogrešek, analiza grobih pogreškov, geodetska opazovanja, geodetska mreža, klasične analize, robustne metode

## KEY WORDS

outlier, outlier analysis, geodetic observations, geodetic network, conventional tests, robust methods

## 1 INTRODUCTION

For the purpose of solving many tasks in geodesy, it is necessary to design and establish a geodetic network of appropriate quality. Due to the influence of various factors, such as careless operators, atmospheric conditions and other factors, observations may contain outliers. Since outliers affect the estimation of unknown parameters and their accuracy in the adjustment of a geodetic network, it is possible to reach wrong conclusions on the basis of the obtained results (Hampel et al., 1986). Therefore, observations must be freed of the influence of outliers. Two most common methods for detecting outliers in geodetic observations are Conventional Outlier Detection Test Procedures and Robust Estimation Methods.

The conventional tests of detecting outliers include Data snooping test (Baarda, 1968) and  $\tau$  test (Pope, 1976). These tests are based on the assumption that observations contain only one outlier. Berber and Hekimoglu have shown that by applying the Data snooping test one outlier can be reliably detected, while if the  $\tau$  test is applied, an outlier cannot be always reliably detected (Berber and Hekimoglu, 2003). Baselga showed that the  $\tau$  test becomes inefficient when several outliers exist in observations (Baselga, 2007). Also, that study showed that the  $\tau$  test is not applicable in cases when observations are correlated. In Saka (2016) two approaches for outlier detection in the Global Navigation Satellite System (GNSS) networks by applying the  $\tau$  test were considered. In the first case one GNSS baseline components were analysed individually and in the second case they were analysed as a whole. When an outlier exists in all baseline components, the second approach is a more efficient solution while the first approach is more efficient in cases when outlier exists in only one baseline component (Saka, 2016). It is asserted and proven in practice, that in case of only one outlier, Conventional Tests are very efficient (Xu, 2005; Hekimoglu et al., 2014).

The robust methods are very suitable statistical tools used for reducing or removing the influence of outliers in observations. In 1964, Huber introduced a robust M-estimation, representing the generalized form of maximum likelihood estimation (Huber, 1964). The robust M-estimation is a very suitable technique for outlier detection in geodetic observations. The principle of the robust-M-estimation application is based on the iterative Least squares (LS) adjustment respecting the condition about gradual changes of the weights of individual observations (Třasák and Štroner, 2014). The LS is very sensitive to deviations from the model assumptions (Hampel et al., 1986) and it extends the effects of outliers to residuals of all observations (Hekimoglu et al., 2011, 2014; Hekimoglu and Erdogan, 2012; Erdogan, 2014). LS play an essential role in all methods of detecting outliers mentioned above. Unlike the LS estimation, the robust M-estimation is far less sensitive to the deviations from the assumed model, hence it resists, to a certain extent, the effect of outliers (Hampel et al., 1986; Třasák and Štroner, 2014).

In order to investigate the effectiveness of outlier detection methods in geodetic observations, Hekimoglu and Koch proposed that as a measure of efficiency the mean success rate (MSR) should be used (Hekimoglu and Koch, 1999, 2000). MSR is given as the number of successes divided by the number of experiments (Hekimoglu and Koch, 1999, 2000). MSR is used for the efficiency measurements of outlier detection methods (Hekimoglu and Erenoglu, 2007; Erenoglu and Hekimoglu, 2010; Hekimoglu et al., 2011, 2014; Erdogan, 2014) and methods of deformation analysis (Hekimoglu, Erdogan and Butterworth, 2010; Nowel, 2016; Sušić et al., 2017).

Below are briefly presented results of researches carried out in cited papers. The robust methods of outlier detection are more efficient than the conventional tests. By increasing the outlier magnitude, the

efficiency of an outlier detection method is also increased. The efficiency of mentioned methods is significantly reduced with the increase of the number of outliers, and the number of unknown parameters. Outlier detection methods are significantly more efficient in cases when the *a priori* standard deviation is unknown. The MSR efficiency indicator reaches a maximum value if there is only one outlier or when the number of unknown parameters is one.

Each observation in geodetic network is measured at least two times, after which its mean value is determined. The obtained mean values are used in the adjustment of geodetic network. When observations are not contaminated by outliers, the mean value represents the optimum value. However, if observations are contaminated by at least one outlier, the mean value will also be contaminated by an outlier. For instance, when observations in a geodetic network are measured twice, the mean value is contaminated by half of an outlier. Since the outlier detection of small magnitudes is much more difficult, it is necessary that the outlier detection is based on original observations, not on their mean values (Erdogan, 2014; Hekimoglu et al., 2014). In his study (Erdogan, 2014), Erdogan has shown that this approach provides higher efficiency of the outlier detection methods. In Hekimoglu et al. (2014) the authors propose the Univariate Approach for the outlier detection. In this approach the analysis is based on a univariate sample which is formed as the difference of original observations measured at least twice. If this approach is used, the efficiency of applied methods increases (Hekimoglu et al., 2014).

## 2 CONVENTIONAL TESTS

The conventional tests of outlier detection are based on local statistical tests (Baarda, 1968; Pope, 1976). When an observation  $l_i$  is contaminated by an outlier  $\delta_i$ , the following hypotheses are tested:

$$H_0: E(\delta_i) = 0 \text{ versus } H_a: E(\delta_i) \neq 0 \tag{1}$$

Hypotheses (1) are tested by applying the Data snooping or  $\tau$  tests (Baarda, 1968; Pope, 1976). The differences between these tests are reflected in the use of different variance factors for the standardization of residuals. In case of the Data snooping test *a priori* standard deviation  $\sigma_0$  is used, and in the  $\tau$  test a posteriori standard deviation  $\hat{\sigma}_0$ .

For the Data snooping test, the test statistics is formed in the following way (Baarda, 1968; Caspary, 2000):

$$w_i = \frac{|v_i|}{\sigma_0 \sqrt{q_{vvi}}} \sim N(0,1), i \in \{1,2,\dots,n\} \tag{2}$$

where  $v_i$  is the residual of the  $i$ -th observation,  $\sigma_0$  a priori standard deviation,  $q_{vvi}$   $i$ -th diagonal element of the cofactor matrix of residuals  $\mathbf{Q}_v$  and  $N$  the standardised Normal distribution. If  $w_i > N_{1-\alpha/2}$ , then  $i$ -th observation is considered as a bad observation. For the significance level  $\alpha$  the value 0.001 is generally chosen, but one often uses the value 0.01 when a need for more rigidity exist.

In the case of the  $\tau$  test, the test statistics is formed in the following way (Pope, 1976; Caspary, 2000):

$$w_i = \frac{|v_i|}{\hat{\sigma}_0 \sqrt{q_{vvi}}} \sim \tau_f, i \in \{1,2,\dots,n\} \tag{3}$$

where  $\hat{\sigma}_0$  is a posteriori standard deviation,  $\tau_f$  is the  $\tau$  distribution with  $f$  degrees of freedom. The  $\tau$  distribution can be expressed as a function of the Student  $t$  distribution:

$$\tau_f = \sqrt{\frac{f t_{f-1}^2}{f - 1 + t_{f-1}^2}} \tag{4}$$

If  $w_i > \tau_{1-\alpha_0/2, f}$ , then the  $i$ -th observation is contaminated by an outlier. If the significance level  $\alpha$  corresponds to all observations, then the significance level  $\alpha_0$  for every individual observation is determined according to the following expression (Pope, 1976; Caspary, 2000):

$$\alpha_0 = 1 - (1 - \alpha)^{1/n} \cong \alpha/n \tag{5}$$

where  $n$  is a number of observations. For the significance level  $\alpha$  the value is generally chosen as  $\alpha = 0.05$ .

As already mentioned, these methods are based on the assumption that only one observation is contaminated by an outlier. If there are more outliers in observations, the theory is inadequate. However, in practice these methods are applied iteratively. Only the observation with the highest test statistic value is tested in one cycle of the iterations. If this observation is rejected, it is removed, and the remaining observations are adjusted again. The procedure is repeated until all observations which are contaminated by outliers are detected. More information about these methods can be found in the following papers Baarda (1968), Pope (1976), Caspary (2000).

### 3 ROBUST METHODS

The robust M-estimation method finds a wide application in detection of outliers in geodetic observations. The principle of the robust M-estimator application is based on an iterative least squares adjustment respecting the condition of a gradual change in weights of individual observations (Koch, 1999):

$$\left. \begin{aligned} \hat{\mathbf{x}}_k &= (\mathbf{A}^T \mathbf{P} \mathbf{W}_{k-1} \mathbf{A})^+ \mathbf{A}^T \mathbf{P} \mathbf{W}_{k-1} \mathbf{l} \\ \mathbf{v}_k &= \mathbf{A} \hat{\mathbf{x}}_k - \mathbf{l} \end{aligned} \right\}_{k=1,2,\dots} \tag{6}$$

where  $\mathbf{A}$  is the design matrix,  $\mathbf{P}$  matrix of observation weights,  $\mathbf{l}$  vector of observations,  $\mathbf{W}$  matrix of robust weights,  $\mathbf{v}$  vector of residuals,  $\hat{\mathbf{x}}$  vector of unknown parameters and  $k$  the number of iterations. The matrix of robust changes in weights of observations  $\mathbf{W}$  may be expressed in the form:

$$\mathbf{W} = \text{diag}(w_1, w_2, \dots, w_n) \tag{7}$$

In the initial iterative step, this matrix is set up as the identity matrix ( $\mathbf{W} = \mathbf{I}$ ). In subsequent iterations, any weight function from the robust M-estimation class can be used for forming robust weights  $w_i$ . In Table 1 the following weight functions are shown: L1 norm (Barrodale and Roberts, 1974), Huber (Huber, 1981), Danish (Krarup, Juhl and Kubik, 1980) and Andrews (Andrews, 1974). Iterative process (6) is performed until all the differences between residuals contained in the vector  $\mathbf{v}_k$  from the current and the corresponding ones contained in the vector  $\mathbf{v}_{k-1}$  related to the preceding iteration, become smaller than the related specified tolerance values for  $\gamma$  (see chapter 5, paragraph 3).

Table 1: Functions of weights.

M-Estimation	Weight Function ( $i \in \{1, 2, \dots, n\}$ )	Critical Value
L1-norm	$w_i = 1/ v_i $	
Huber	$w_i = \begin{cases} 1, &  v_i  \leq c \\ c/ v_i , &  v_i  > c \end{cases}$	$c \in [1.5\sigma_0, 2\sigma_0]$
Danish	$w_i = \begin{cases} 1, &  v_i  \leq c \\ \exp(- v_i /c), &  v_i  > c \end{cases}$	$c \in [1.5\sigma_0, 2\sigma_0]$
Andrews	$w_i = \begin{cases} \sin( v_i /c) / ( v_i /c), &  v_i  \leq c\pi \\ 0, &  v_i  > c\pi \end{cases}$	$c \in [1.5\sigma_0, 2\sigma_0]$

The principle of outlier detection is based on the comparison of estimated observation residuals from robust adjustment with critical value. When the estimated observation residuals value from the last iteration exceeds the critical value  $3\sigma_0$ , the observation is considered as a bad observation (Hekimoglu et al., 2014). When *a priori* standard deviation  $\sigma_0$  is unknown, the critical value is set to  $3\hat{\sigma}_0$ , where  $\hat{\sigma}_0$  is *a posteriori* standard deviation from the first iteration. A more detailed information about the robust methods can be found in Huber (1981), Hampel et al. (1986), Koch (1999), Trásák and Štroner (2014).

#### 4 EFFICIENCY OF OUTLIER DETECTION METHODS

The outlier detection in observations is considered successful if all outliers in observations were discovered using the appropriate method. The efficiency of an outlier detection method in observations cannot be determined on the basis of analysing only one set of real observations in a geodetic network. In this case it is not known which observations contain outliers so it is not possible to check the results obtained using the appropriate method. Also, the efficiency analysis refers to only one of many different models of geodetic networks, and therefore it is not known how methods behave in other cases.

It has been already mentioned that the efficiency of the outlier detection method in geodetic networks can be determined by using the MSR factor (Hekimoglu and Koch, 1999, 2000). MSR is a very good empirical measure of the efficiency of the outlier detection methods. It is computed on the basis of a large number of simulated observation sets. MSR is calculated as the ratio of the number of observation sets in which outliers were correctly detected and the total number of simulated observation sets (Hekimoglu and Koch, 1999, 2000; Nowel, 2016).

In practice geodetic observations contain Normally distributed random errors. Taking into account the assumption that observations are not correlated, it is natural that random errors of observation have the following form:

$$\varepsilon_i \sim N(0, \sigma_i^2), i \in \{1, 2, \dots, n\} \tag{8}$$

where  $\sigma_i$  is standard deviation of the *i*-th observation. Random observation errors can be generated by applying a random generator. In the simulation signs and locations of outliers are chosen randomly. The magnitudes of outliers  $\delta_i$  take values from the interval:

$$\delta_i \in [a\sigma_0, b\sigma_0] \quad (9)$$

where  $\sigma_0$  is *a priori* standard deviation,  $a \geq 1$  defines the lower limit of the interval, whereas  $b > 1$  defines the upper one (Hekimoglu and Erenoglu, 2007). The values from this interval are Uniformly distributed.

The vector of simulated observations is formed in the following manner:

$$\mathbf{l}_{obs} = \mathbf{l} + \boldsymbol{\varepsilon} + \boldsymbol{\delta} \quad (10)$$

where  $\mathbf{l}$  is the vector of *theoretical observations* (it can be determined on the basis of approximate coordinates of points),  $\boldsymbol{\varepsilon}$  vector of random observation errors and  $\boldsymbol{\delta}$  is the vector of outliers. It is necessary to form independently a large number of sets of simulated observations according to expression (10). The detection of outliers in observations, by applying the appropriate method, is carried out for every set of simulated observations.

MSR is computed in the following way (Hekimoglu and Koch, 1999, 2000; Nowel, 2016):

$$MSR(n_o) = S/N \quad (11)$$

where  $n_o$  is the number of outliers in observations,  $S$  is the number of sets of simulated observations wherein outliers are correctly detected and  $N$  is the total number of simulated observation sets. In order to make the comprehensive analysis it is necessary to determine the MSRs for various values of  $n_o$ . More detailed explanations concerning the MSR factor can be found, for example, in Hekimoglu and Koch (1999, 2000), Hekimoglu and Erenoglu (2007), Erenoglu and Hekimoglu (2010), Hekimoglu, Erdogan and Butterworth (2010), Nowel (2016).

## 5 EXPERIMENTAL RESEARCH

The subject of the analysis in this paper is the Pobednik statue, which is located at the Kalemegdan fortress in Belgrade. The Pobednik statue is a male sculpture made of bronze, placed on a construction consisting of a stone base and a raised stone pillar. By visual inspection of the construction of the monument, certain deformations were observed on the construction itself. For the purpose of monitoring condition of the construction of the statue, geodetic control network which consists of five points was designed and established. For the monitoring the state of the monument's construction since 2008, geodetic deformation measurements from the control geodetic network are periodically carried out. In the network directions in two repetitions were measured with standard deviation  $\sigma_p = 2''$  and distances in two repetitions with standard deviation  $\sigma_d = 2\text{mm} + 2\text{ppm}$ . Fig. 1 presents position of the points as well as distances and directions. The procedure of outlier detection in the geodetic network is carried out using the Data Snooping test. However, sometimes if outliers in observations are not detected and eliminated, it may affect the objectivity of the procedure deformation analysis. Thus, the correct detection of outliers in observations is a prerequisite for the effective application of the deformation analysis method. The analysis of the efficiency of different approaches for outlier detection in geodetic networks was performed below, on the example of the control network for geodetic monitoring of the Pobednik statue. In this study, the subject of the analysis is a two-dimensional triangulation-trilateration network for geodetic monitoring, while the one-dimensional levelling networks are most frequently analysed in the existing literature (Hekimoglu et al. 2014; Erdogan, 2014).

Observations are simulated with random errors which are Normally distributed with adopted standard deviation  $\sigma = 2''$  for directions, and  $\sigma = 2\text{mm} + 2\text{ppm}$  for distances. In the experiment two variants with different interval of magnitudes are considered. In the first variant the magnitudes of outliers take values from the interval  $[3\sigma_0, 6\sigma_0]$ , whereas in the other one the interval is  $[6\sigma_0, 12\sigma_0]$ . For both variants following four cases are treated: (1) observations contain no outliers ( $n_o = 0$ ); (2) one randomly selected observation is contaminated by an outlier ( $n_o = 1$ ); (3) two randomly selected observations are contaminated by outliers ( $n_o = 2$ ) and (4) three randomly selected observations are contaminated by outliers ( $n_o = 3$ ). Five thousand sets of simulated observations were generated (first and second repetitions) for each of the four mentioned cases for both variants, according to equation (10). The simulations were performed by applying the Monte Carlo method using the software package Matlab R2013a.

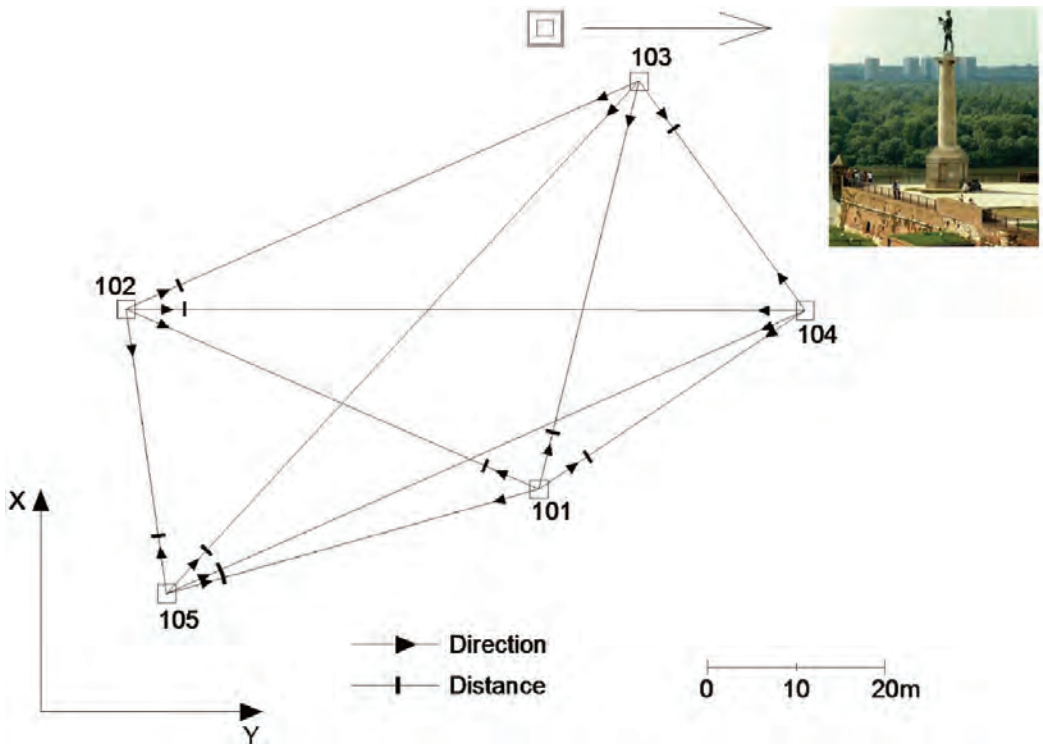


Figure 1: Geodetic network for Pobednik statue in the Kalemegdan Fortress in Belgrade.

Outliers are analysed for every set of simulated observations by applying the classical and new approaches to the outlier analysis. The classical approach involves using the mean values of observations from two repetitions for the outlier analysis. The new approach by Hekimoglu et al. (2014) and Erdogan (2014) involves using the original observations for the outlier analysis. Outliers are detected by applying the conventional tests for outliers and the robust methods. For the significance level  $\alpha$  the following values chosen as 0.001 and 0.05, for the Data snooping and the  $\tau$  test respectively. In Huber, Danish and Andrews methods for  $c$  the assumed value is chosen as  $1.5\sigma_0$ . In the case of the robust methods for the tolerance  $\gamma$  the assumed value is chosen as  $0.1''$  for directions and  $0.1\text{ mm}$  for distances. The MSR efficiency coefficients are calculated independently for each method in both approaches (Table 2 and Table



3). The comparison analysis of the MSRs, for both individual approaches and individual methods, for different outlier magnitudes intervals  $[3\sigma_0, 6\sigma_0]$  and  $[6\sigma_0, 12\sigma_0]$ , is presented in Figs. 2 and 3, respectively.

Table 2: The MSRs [%] for conventional tests and robust methods, classical approach.

Method	$n_o = 0$	$\delta_i \in [3\sigma_0, 6\sigma_0]$			$\delta_i \in [6\sigma_0, 12\sigma_0]$		
		$n_o = 1$	$n_o = 2$	$n_o = 3$	$n_o = 1$	$n_o = 2$	$n_o = 3$
DS test	99.96	1.16	0.30	0.00	49.56	24.34	10.48
$\tau$ test	95.36	14.06	1.50	0.26	69.64	33.82	9.32
L1 norm	99.08	6.64	0.72	0.10	65.28	43.72	28.66
Huber	99.88	4.02	0.50	0.04	54.32	30.72	16.70
Danish	99.00	9.46	1.44	0.20	73.98	53.80	36.18
Andrews	99.88	4.16	0.50	0.00	54.54	29.58	14.76

Table 3: The MSRs [%] for conventional tests and robust methods, new approach.

Method	$n_o = 0$	$\delta_i \in [3\sigma_0, 6\sigma_0]$			$\delta_i \in [6\sigma_0, 12\sigma_0]$		
		$n_o = 1$	$n_o = 2$	$n_o = 3$	$n_o = 1$	$n_o = 2$	$n_o = 3$
DS test	94.14	60.70	40.10	23.64	94.46	93.52	91.62
$\tau$ test	94.86	51.54	26.80	10.74	94.74	92.96	89.40
L1 norm	81.18	55.38	41.58	29.98	73.70	67.72	61.50
Huber	86.80	61.68	44.96	30.52	85.10	80.50	74.84
Danish	67.96	53.30	43.66	35.80	64.98	62.98	61.16
Andrews	86.74	62.74	45.02	30.66	86.96	84.70	83.64

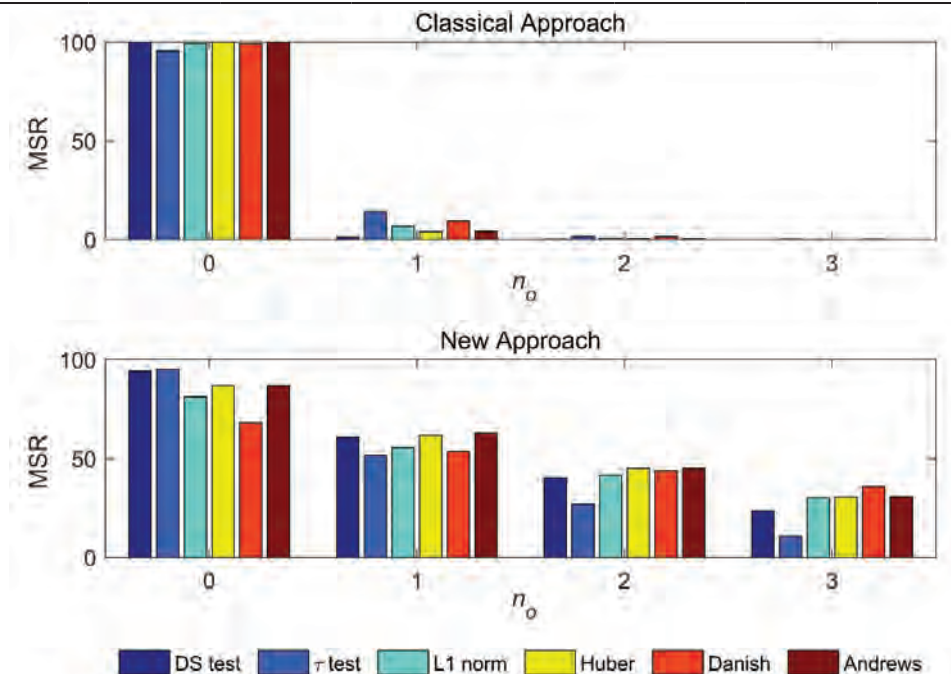


Figure 2: The MSRs of conventional tests and robust methods,  $\delta_i \in [3\sigma_0, 6\sigma_0]$ .



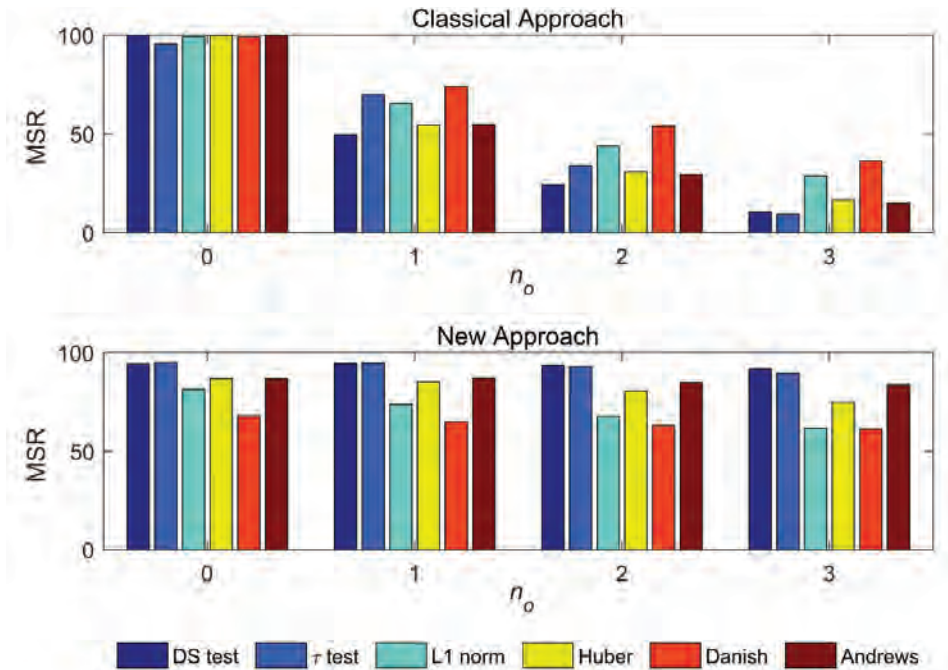


Figure 3: The MSR for conventional tests and robust methods,  $\delta_i \in [6\sigma_o, 12\sigma_o]$ .

Based on previous research, it is known that the efficiency of different outlier detection in observations, among other things, is influenced by random measurement errors, the number of unknown parameters, the magnitude of outliers, and the type and configuration of the geodetic network (Hekimoglu and Erenoglu, 2007; Erenoglu and Hekimoglu, 2010; Hekimoglu et al., 2011, 2011a, 2014; Erdogan, 2014). For this reason some observations which are not contaminated by outliers are detected as observations with outliers (first-type error ( $\alpha$ )), while observations which are contaminated by outliers are detected as those without outliers (second-type error ( $\beta$ )). As it has been previously established efficiency of the outlier detection methods for geodetic networks is different. Below is the analysis of the influence of the first type error ( $\alpha$ ) and the second type error of the ( $\beta$ ) on the above mentioned differences. The analogous analysis concerning the deformation-analysis methods is available in Nowel (2016).

Within the experiment only two observations are analysed, observations of direction  $\alpha_{102-103}$  and direction  $\alpha_{105-102}$ . The assumption is that observation of direction  $\alpha_{102-103}$  from the first repetition is contaminated by outlier, while observation of direction  $\alpha_{105-102}$  from the first repetition is not. It is also assumed that one more randomly selected observation is contaminated by an outlier. The magnitudes of outliers take values from  $\delta_i \in [3\sigma_o, 6\sigma_o]$ . As in the previous experiment, five thousand sets of simulated observations have been generated.

The outlier detection is performed for every set of simulated observations by applying the mentioned approaches, where the conventional tests and the robust methods are applied. For the significance level  $\alpha$  and critical value  $c$  the same values were chosen as in the previous experiment. The number of cases wherein the direction  $\alpha_{105-102}$  is detected as an observation with an outlier (first-type error ( $\alpha$ )) is

determined, as well as the number of cases wherein the direction  $\alpha_{102-103}$  is detected as an observation without an outlier (second-type error ( $\beta$ )). The first-type and second-type errors for each method for both approaches are presented in Table 4. The histograms of the test statistics  $w_{\alpha_{105-102}}$  and  $w_{\alpha_{102-103}}$  for the Data snooping test and the magnitude of residuals  $v_{\alpha_{105-102}}$  and  $v_{\alpha_{102-103}}$  for the Danish method are presented in Figs. 4 and 5.

Table 4: First-type and second-type errors for conventional tests and robust methods.

Method	Classical approach		New approach	
	$\alpha$ [%]	$\beta$ [%]	$\alpha$ [%]	$\beta$ [%]
DS test	0.00	99.24	0.16	45.88
$\tau$ test	0.30	94.92	0.08	63.08
L1 norm	0.34	95.80	0.62	39.12
Huber	0.00	99.38	0.26	40.94
Danish	0.32	90.74	0.96	27.32
Andrews	0.00	99.36	0.24	41.58

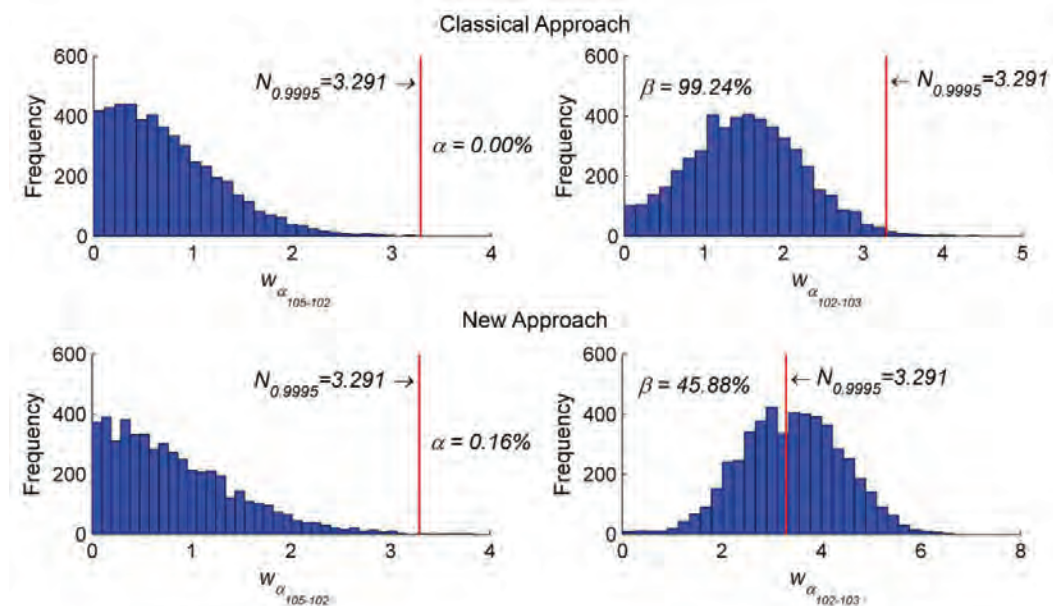


Figure 4: Histograms of test statistics  $w_{\alpha_{105-102}}$  and  $w_{\alpha_{102-103}}$  for Data snooping test.

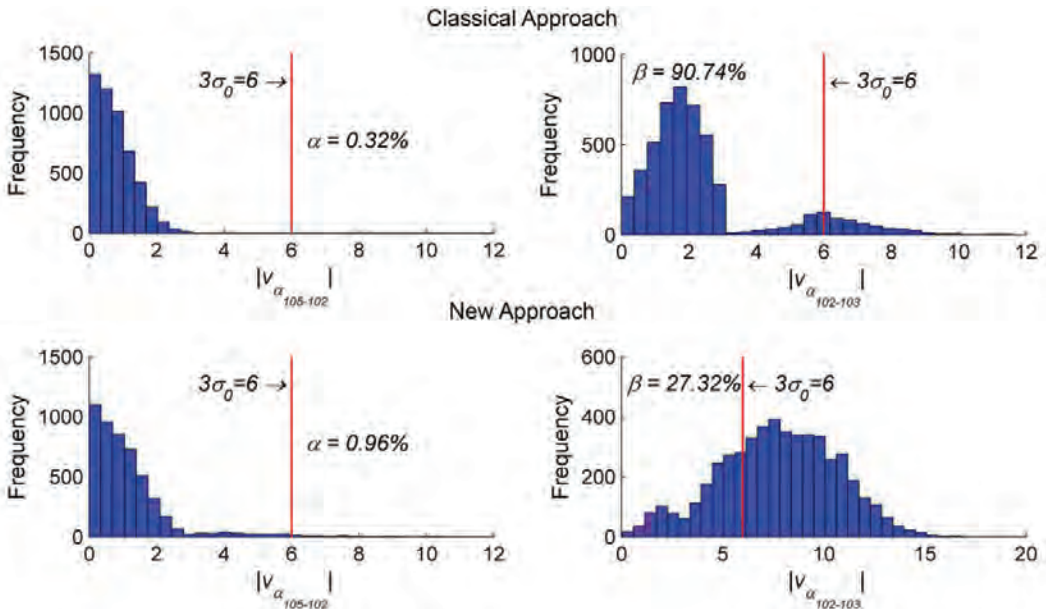


Figure 5: Histograms of magnitude of residuals  $v_{\alpha_{105-102}}$  and  $v_{\alpha_{102-103}}$  for Danish method.

## 6 DISCUSSION AND CONCLUSION

The efficiency analysis of different approaches for outlier detection in geodetic networks has been conducted on the example of the control network of geodetic monitoring for the Pobednik statue in the Kalemegdan Fortress in Belgrade. The outlier detection is performed by applying the conventional tests and the robust methods. MSR factor was used for measuring efficiency in this study. The MSR of the applied methods for the classical approach are presented in Table 2. When observations are not contaminated by outliers ( $n_o = 0$ ), the MSR has large values, which indicates that the mean value is an optimal estimate when observations are not contaminated by outliers. When observations are contaminated by outliers, the MSR is significantly reduced, because when calculating the mean value, the magnitude of the outlier is reduced. The MSR of the conventional tests and the robust methods for the new approach are presented in Table 3. The efficiency of the applied methods in this case is significantly higher than in the case of the classical approach. The efficiency differences between different approaches for outlier analysis and individual outlier detection methods are easily noticeable in Figs. 2 and 3.

In addition to the efficiency analysis of different approaches in the outlier detection in this paper, an analysis was made to determine which error (first-type and second-type errors ( $\alpha$  and  $\beta$ )) affect these efficiency differences more. The first-type and second-type errors of the conventional tests and the robust methods are presented in Table 4. The first-type errors of the applied methods are very similar for both approaches, whereas the second-type errors are different significantly. The differences between the second-type errors for the Data snooping test and the Danish method for individual approaches to the outlier detection can be noticed in Figs. 4 and 5. The minimal values of the second-type errors for the applied methods concern the new approach to the analysis of outliers. For the new approach the smallest values

of the second-type errors occur for the Danish (27.32%) method, whereas the largest one occurs for the  $\tau$  test (63.08%). Thus, the efficiency differences between individual methods for detecting outliers are noticeable when observations are contaminated by outliers, while the efficiency of detecting observations which are not contaminated by outliers is very similar.

In the case of Example (2), one randomly selected observation contaminated by outlier was selected. By applying a new approach to the outlier detection, the efficiency of the Data snooping test was increased by 59.54% and the Huber method's efficiency by 57.66% when the magnitude of the outliers took values from the interval  $[3\sigma_0, 6\sigma_0]$ . If the magnitude of the outliers takes values from the interval  $[6\sigma_0, 12\sigma_0]$ , the Data snooping test efficiency is increased by 44.9% while the Huber method's efficiency is increased by 30.78%. Erdogan (2014) obtained similar results, the efficiency of the Data Snooping test and the Huber method increased by 77.6% and 79.1%, respectively, when the magnitude of the outliers takes values from the interval  $[3\sigma_0, 6\sigma_0]$ . When the magnitude of the outliers takes values from the interval  $[6\sigma_0, 12\sigma_0]$ , the efficiency of the Data Snooping test is increased by 35.1% and the Huber method by 24.0%. The obtained results have shown that the new approach for the outlier detection ensures greater efficiency of the conventional tests and the robust methods, which is analogous to conclusions published in Hekimoglu et al. (2014) and Erdogan (2014).

## Acknowledgments

The authors gratefully acknowledge the funding provided by the research project of the Department of Civil Engineering and Geodesy (DGG) of the Faculty of Technical Sciences in Novi Sad in 2017.: "Improvement of educational process and theoretical and applied research in civil engineering".

## Literature and references:

- Andrews, D.F. (1974). A Robust Method for Multiple Linear Regression. *Technometrics*, 16 (4), 523–531. DOI: <http://dx.doi.org/10.1080/00401706.1974.10489233>
- Baarda, W. (1968). A testing procedure for use in geodetic networks. Delft: Netherlands Geodetic Commission.
- Barrodale, I., Roberts, FDK. (1974). Solution of an overdetermined system of equations in the L1 norm. *Commun ACM*, 17 (6), 319–320. DOI: <http://dx.doi.org/10.1145/355616.361024>
- Baselga, S. (2007). Critical Limitation in Use of Test for Gross Error Detection. *J Surv Eng*, 133 (2), 52–55. DOI: [http://dx.doi.org/10.1061/\(ASCE\)0733-9453\(2007\)133:2\(52\)](http://dx.doi.org/10.1061/(ASCE)0733-9453(2007)133:2(52))
- Berber, M., Hekimoglu, S. (2003). What Is the Reliability of Conventional Outlier Detection. *Surv Rev*, 37 (290), 308–318. DOI: <http://dx.doi.org/10.1179/sre.2003.37.290.308>
- Caspary, W.F. (2000). Concepts of Network and Deformation Analysis. Kensington: The University of New South Wales, School of geomatic engineering.
- Erdogan, B. (2014). An Outlier Detection Method in Geodetic Networks Based on the Original Observations. *Bol Ciências Geodésicas*, 20 (3), 578–589. DOI: <http://dx.doi.org/10.1590/S1982-21702014000300033>
- Erenoglu, R.C., Hekimoglu S (2010). Efficiency of robust methods and tests for outliers for geodetic adjustment models. *Acta Geod Geophys Hungarica*, 45 (4), 426–439. DOI: <http://dx.doi.org/10.1556/AGeod.45.2010.4.3>
- Hampel, F., Ronchetti, E., Rousseeuw, P., Stahel, W. (1986). *Robust statistics: The approach based on influence function*. New York: Wiley.
- Hekimoglu, S., Erdogan, B. (2012). New Median Approach to Define Configuration Weakness of Deformation Networks. *J Surv Eng*, 138 (3), 101–108. DOI: [http://dx.doi.org/10.1061/\(ASCE\)SU.1943-5428.0000080](http://dx.doi.org/10.1061/(ASCE)SU.1943-5428.0000080)
- Hekimoglu, S., Erdogan, B., Butterworth, S. (2010). Increasing the Efficacy of the Conventional Deformation Analysis Methods: Alternative Strategy. *J Surv Eng*, 136 (2), 53–62. DOI: [http://dx.doi.org/10.1061/\(ASCE\)SU.1943-5428.0000018](http://dx.doi.org/10.1061/(ASCE)SU.1943-5428.0000018)
- Hekimoglu, S., Erdogan, B., Erenoglu, R.C., Hosbas, R.G. (2011). Increasing the efficacy of the tests for outliers for geodetic networks. *Acta Geod Geophys Hungarica*, 46 (3), 291–308. DOI: <http://dx.doi.org/10.1556/AGeod.46.2011.3.2>
- Hekimoglu, S., Erenoglu, R.C., Sanli, D.U., Erdogan, B. (2011a). Detecting configuration weaknesses in geodetic networks. *Survey Review*, 43 (323), 713–730. DOI: <https://doi.org/10.1179/003962611X13117748892632>
- Hekimoglu, S., Erdogan, B., Soykan, M., Durdag, U.M. (2014). Univariate Approach for Detecting Outliers in Geodetic Networks. *J Surv Eng*, 140 (2), 1–8. DOI: [http://dx.doi.org/10.1061/\(ASCE\)SU.1943-5428.0000123](http://dx.doi.org/10.1061/(ASCE)SU.1943-5428.0000123)
- Hekimoglu, S., Erenoglu, R.C. (2007). Effect of heteroscedasticity and heterogeneity on outlier detection for geodetic networks. *J Geod*, 81 (2), 137–148. DOI: <http://dx.doi.org/10.1007/s10062-007-0007-0>

dx.doi.org/10.1007/s00190-006-0095-z

Hekimoglu, S., Koch, K. (1999). How can reliability of the robust methods be measured. In: M. Altan (Ed.), Gründige L (eds) Third Turkish-German joint geodetic days (pp. 179–196). Istanbul: Istanbul Technical University.

Hekimoglu, S., Koch, K. (2000). How can reliability of the test for outliers be measured. *Allg Vermessungs-Nachrichten*, 107 (7), 247–254.

Huber, P.J. (1964). Robust Estimation of a Location Parameter. *Ann Math Stat*, 35 (1), 73–101. DOI: <http://dx.doi.org/10.1214/aoms/1177703732>

Huber, P.J. (1981). *Robust Statistics*. New York: Jhon Wiley & Sons.

Koch, K-R. (1999). *Parameter Estimation and Hypothesis Testing in Linear Models*. Berlin: Springer Science & Business Media.

Krarup, T., Juhl, J., Kubik, K. (1980). Götterdämmerung over least squares adjustment. In: 14th congress of international society photogrammetry (pp. 369–378). Hamburg.

Nowel, K. (2016). Investigating efficacy of robust M-estimation of deformation from observation differences. *Surv Rev*, 48 (346), 21–30. DOI: <http://dx.doi.org/10.1080/00396265.2015.1097585>

Pope, A.J. (1976). *The statistics of residuals and the detection of outliers*. U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Ocean Survey, Geodetic Research and Development Laboratory.

Saka, M.H. (2016). Performance evaluation of outlier detection methods in GNSS vector networks using 1D and 3D component analysis. *Meas J Int Meas Confed*, 82 (1), 145–150. DOI: <http://dx.doi.org/10.1016/j.measurement.2015.12.029>

Sušić, Z., Batilović, M., Ninkov, T., Bulatović, V., Aleksić, I., Nikolić, G. (2017). Geometric deformation analysis in free geodetic networks: case study for Fruška gora in Serbia. *Acta Geodyn Geomater*, 14 (3), 341–355. DOI: <http://dx.doi.org/10.13168/AGG.2017.0017>

Tršák, P., Štroner, M. (2014). Outlier detection efficiency in the high precision geodetic network adjustment. *Acta Geod Geophys*, 49 (2), 161–175. DOI: <http://dx.doi.org/10.1007/s40328-014-0045-9>

Xu, P. (2005). Sign-constrained robust least squares, subjective breakdown point and the effect of weights of observations on robustness. *J Geod*, 79 (1), 146–159. DOI: <http://dx.doi.org/10.1007/s00190-005-0454-1>



Batilović M., Sušić Z., Marković M. Z., Vujinović M., Nikolić G., Ninkov T. (2018). The efficiency of different outlier detection approaches in geodetic networks: case study for Pobednik statue. *Geodetski vestnik*, 62 (2), 293-305. DOI: 10.15292/geodetski-vestnik.2018.02.293-305

**Mehmed Batilović, M.Sc.**

University of Novi Sad, Faculty of Technical Sciences  
Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia  
e-mail: mehmed@uns.ac.rs

**Marijana Vujinović, M.Sc.**

University of Novi Sad, Faculty of Technical Sciences  
Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia  
e-mail: marijana.petkovic@uns.ac.rs

**Assist. Prof. Zoran Sušić, Ph.D.**

University of Novi Sad, Faculty of Technical Sciences  
Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia  
e-mail: zsusic@uns.ac.rs

**Assist. Prof. Gojko Nikolić, Ph.D.**

University of Montenegro, Institute for Geography  
Danila Bojovića 3, 81400 Nikšić, Montenegro  
e-mail: gojkor@ac.me

**Marko Marković, M.Sc.**

University of Novi Sad, Faculty of Technical Sciences  
Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia  
e-mail: marko\_m@uns.ac.rs

**Prof. Toša Ninkov, Ph.D.**

University of Novi Sad, Faculty of Technical Sciences  
Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia  
e-mail: ninkov.tosa@gmail.com