



UDK 528.1

## COMPARISON OF OUTLIER DETECTION AT THE EDGES OF POINT CLOUDS USING STATISTICAL APPROACH AND FUZZY METHODOLOGY: GROUND-BASED LASER SCANNER FIELD EXPERIMENT AND RANDOMLY SIMULATED POINT CLOUD

Reza ARABSHEIBANI<sup>1</sup>, Abbas ABEDINI<sup>2</sup>, Yousef KANANI SADAT<sup>3</sup>

*Department of Surveying and Geomatics Engineering, College of Engineering, University of Tehran, North Kargar Ave., Jalal-Al-Ahmad Crossing, Tehran, Iran*

*E-mails: <sup>1</sup>rasheibani@ut.ac.ir; <sup>2</sup>aabedeni@ut.ac.ir; <sup>3</sup>yousefkanani@ut.ac.ir (corresponding author)*

*Received 27 March 2015; accepted 15 August 2015*

**Abstract.** The random error is following the features of normal distribution function (NDF) which those random errors deviated from the NDF's characteristics can be considered as outliers. In fact, the outliers exist inevitably in any observed parameter that is an undesirable part of the measurement's procedure due to its negative influence on the sensitivity analysis. It is therefore necessary to investigate more efficient methodologies especially for current remote sensing data processing and assimilations. In this paper, the comparisons of Baarda method as the conventional statistical methodology with the Fuzzy approach are presented to detect the outliers at the edges of two data groups: 1. The point cloud of ground-based laser scanner field experiment from one side of a wall, and 2. A group of randomly simulated distributed 3D point cloud. The results show that the Baarda method eliminates the outliers as soon as they are being found while the Fuzzy approach works critically based on the outputs of the statistical tests. Thus, the Fuzzy approach deals mostly with the residuals and those observed errors in the adjustment computational procedures. The obtained results about the successfulness rate of outlier detection for each method are separately presented in both graphical and statistical overview. Also, the capabilities of Fuzzy approach to detect the outliers in different point cloud's size and numbers of existing outliers at the edges of point cloud are investigated and discussed in details.

**Keywords:** fuzzy logic, point cloud, outlier detection, random error, Baarda method, laser scanner.

### Introduction

Nowadays the comprehensive applications of laser scanners especially the ground-based ones necessitate to improve the quality of their products. In general, using the new remote sensing techniques and sensors especially the ground-based ones induce us to assess and revisit the data mining, assimilation and modeling approaches (Najibi, Jin 2013; Najibi, Arabsheibani 2013) by taking into consideration the optimization of time and cost-efficiency. These ground-based laser scanners are able to acquire direct, accurate and time-saving observations in form of three-dimension (3D) points called hereafter as point clouds. The 3D point clouds are taken by the ground-based laser scanner directly over the targets. However, even one deviant observation may lead to destructive divergences during

further procedures, particularly in 3D modeling and texture matching. Therefore, it is necessary to pre-process the 3D point clouds precisely in order to detect the outliers and then eliminate them from the raw observations (Jahandideh *et al.* 2014). Furthermore, the detection of outliers in point cloud is not only a simple process due to lack of information about the point's distribution type, but also the existence of random errors in data set make it more convoluted (Rao, Toutenburg 1995), especially where the distribution of points might be varied or very complicated in different parts of a single object (Sotoodeh 2006).

Practically, those outliers with similar values to the random errors can only be detected by the outlier determination tests (Sisman *et al.* 2012), while the conventional outlier detection methodologies which were

firstly introduced by (Baarda 1968; Baarda, Commission 1967) are looking for the outliers using the iterative procedures (Koch 1999). Notwithstanding, the Robust Estimation (RE) method which was firstly suggested by Huber (2011) and later improved by Hampel *et al.* (2011), and it can be applied as an alternative method instead of the present conventional methodologies of outlier detection.

On the other hand, recently the well-known approach named Artificial Intelligence (AI) is suggested for outlier detection. AI methodology which firstly proposed by (Zadeh 1965) uses the integrated neural networks and Fuzzy logic techniques. Recent studies showed that in addition to the use of Fuzzy method for outlier detection in the geodetic networks, it can be also used for detecting the outliers in the point clouds (Aliosmanoğlu *et al.* 2002; Berberan 1995; Breunig *et al.* 2000; Gökalp *et al.* 2008; Sun 1994).

It is worth mentioning here that by using the common methods of outlier detection, it is possible to detect only one outlier in each solution's step and then this outlier will be withdrawn later from the rest of observations. This procedure continues until it satisfies a specified accuracy criterion. More importantly, although sometimes a valuable observation – in spite of whether it is an outlier or a normal one – can help us to reach more precise and accurate outcome (Monhor *et al.* 2005). In fact, the conventional estimators of statistical approaches for outlier detection are following the logical concept of binary (0 or 1). Accordingly, with respect to the critical values whether those observations having large deviations or located in about of the critical values can be considered completely as the outliers. Conversely, labelling observations as crisp outlier or normal one, that is the major deficiency of conventional methods which is resolved in Fuzzy methodology and thus the percentage of being outlier for an observation is assessed through using subscription function. In addition to this, the procedure of Fuzzy outlier detection is not an iterative one while the iterative calculations through all observations are a major obstacle in all aforementioned conventional methods. The Fuzzy method can work based on the procedures derived from the statistical tests and thus the results of the outlier detection will be even more stable.

In general, many studies have focused on the outlier detection methodologies within the point clouds, but most of them are done for a limited number of observations and a few ones for the edges of point clouds. It is clear that detecting outliers in point cloud needs more additional studies in particular for

the ground-based laser scanners which their users are becoming more and more these days. The discontinuity of those observed points reflected from different sides of an object as well as the inhomogeneous distribution of unwanted noise in the ground-based laser scanner's point cloud make this issue more controversial than before. Based on this, Papadimitriou *et al.* (2003) divided the primary outlier detection approaches as the following categories: distribution based approach, depth based and cluster based approach. Accordingly, the distribution based approaches such as Fuzzy methodology apply statistical models by considering the critical realization level. In this approach, the data distribution's function is already known but generally speaking it is costly for common applications of the point clouds. Conversely, the depth based approach works based on the definition of depth and formation of the points in each layer by considering this fact that the shallow layers holding high probability (in terms of the number of outliers) rather than the deep ones (Johnson *et al.* 1998). In the cluster based approach, the whole data should be clustered and those points which are not involved in any cluster will be considered as the outliers.

In this study, firstly we apply the Baarda method as a statistical methodology and the Fuzzy algorithm to detect the outliers at the edges of point cloud derived from a ground-based laser scanner. Next, these two approaches have been compared through applying onto several sets of randomly distributed point cloud. It is initially assumed here that these data groups are holding the normal distribution function (NDF) properties. Moreover, it is presented and discussed in this paper that the outputs of the common statistical method as well as the adjustment matrices together can be considered as the inputs for the Fuzzy algorithm and in this case, the accuracy of Fuzzy method depends critically on the outputs of these conventional methods.

Section 1 of the paper presents the theoretical definition of Baarda method and Fuzzy approach applied in this work. The case study and data collection field experiment and also the randomly simulated distributed point clouds are given in Section 2. The results and discussions are shown in Section 3 including Baarda method and Fuzzy approach utilization as well as their comparisons and discussions followed by the last section which is about Summary and conclusions.

## 1. Theory and methodologies

### 1.1. Baarda method

The outlier is an observed measurement which has a considerable deviation with respect to a specified

assumed criterion (Muñoz-Garcia *et al.* 1990). In conventional outlier detection methodologies including t-test,  $\tau$ -test and the data snooping (DS), outliers will be eliminated which would cause some undesirable changes in the model's formation. Conventional approaches basically remove only a single blunder during each step of adjustment calculations using least square solution (LSS) and post-processing tests for uncertain observations. LSS can provide us with an unbiased and randomized estimator if a stable functional model is introduced during each adjustment procedure and calculation step.

In Baarda methodology, it is supposed that at least one outlier exists within the entire set of observations. It is not feasible to find more than one outlier in each calculations step of Baarda method, so when the variances of observations are known or computed precisely, the Baarda methodology can be useful. Therefore, it can give us a reasonable prediction for the standard deviations of observations' residuals while the computed residuals will follow the NDF features.

As it is shown in Table 1,  $N$  stands for observation matrix,  $S_0$  is the initial standard deviation based on unite weight,  $f$  is the degree of freedom,  $\alpha_0$  is reference level,  $M$  stands for the normal distribution,  $F$  is the Fisher distribution,  $\chi^2$  is the estimator of  $K^2$ ,  $t$  states the t-student distribution and finally  $\tau$  indicates the estimator of tau distribution. Here, initially and for the sake of calculation's simplicity, it is supposed that there is no correlation between the observations and residuals and thus the  $\alpha_0$  as reference level can be computed as:

$$\alpha_0 = 1 - (1 - \alpha)^{1/n} \cong \frac{\alpha}{n}, \quad (1)$$

where  $n$  is the number of observations and  $\alpha$  is considered typically as 5 or 10 percent.

Table 1. Statistical tests and the relevant critical values (Koch 1999)

Test	Estimator	Critical value
DS	$\frac{(Nv)_i}{s_0 \sqrt{(NQ_{vv}N)_{ii}}}$	$M_{1-\alpha_0/2} = \sqrt{F_{1,\infty,1-\alpha_0}}$ $= \sqrt{\chi^2_{1,\infty,1-\alpha_0}}$
Tau	$\frac{(Nv)_i}{m_0 \sqrt{(NQ_{vv}N)_{ii}}}$	$\tau_{f,1-\alpha_0/2} =$ $\sqrt{\frac{f \times t_{f-1,1-\alpha_0/2}^2}{f-1+t_{f-1,1-\alpha_0/2}^2}}$
t	$\frac{(Nv)_i}{\bar{m}_0 \sqrt{(NQ_{vv}N)_{ii}}}$	$t_{f-1,1-\alpha_0/2}$

Since most of the methods which work based on LSS are not stable (due to lack of stable functional model), a unique and global criteria to weight and estimate the optimized value to compare them is not available either. Therefore, the practical estimators to deal with them are M-Estimators, L-Estimators and E-Estimators (Gökalp *et al.* 2008). The LSS of residuals is a specific section of M-Estimators which its objective function can be defined as follows (Vaníček, Wells 1972):

$$\sum_{i=1}^n P_i v_i^2 \rightarrow \min, \quad (2)$$

and then the procedure to estimate the unknown parameters can be given as:

$$x_i = (A^T \bar{P}_i A)^{-1} A^T \bar{P}_i l; \quad (3)$$

$$v_i = A x_i - l, \quad (4)$$

where in Eq. (2), Eq. (3) and Eq. (4),  $P_i$  is the weighted matrix,  $A$  is the design matrix,  $v$  stands for the residuals vector,  $x$  is the unknown vector,  $l$  is the observed vector and  $i$  is the iteration index in each solution procedure. In above equations,  $P_i$  is being calculated as follows:

$$\bar{P}_i = P W_{i-1}; \quad (5)$$

$$A^T \bar{P} v = 0, \quad (6)$$

where  $W$  will be considered the initial weighting factor in first round of LSS computations which can be given as:

$$W_0 = I_{n \times n}. \quad (7)$$

In Eq. (7),  $n$ -index refers to the total number of observations. It is clear that the corresponding iterations are continuing until the difference between  $x_{i+1}$  and  $x_i$  approaches to a determined value or extremely a low amount. Moreover, it can be seen in final step that the weighted matrix related to the outliers are getting close to zero but the weighted matrix of normal observation is unvaried.

Eq. (1) to Eq. (7) are mostly related to the determination of analytical weighted matrix to make a decision whether an observation is an outlier or a normal one. Conversely, in order to get an analytical weighted matrix corresponding to the stable weighted factor, it is possible to compare the residuals with those critical values derived from the in-hand LSS computations or those real known values. The procedure to get the critical values is shown as follows:

$$C_i = s_0 \sqrt{Q_{VV_{ii}}} \sqrt{P_{ii}} t_{f,1-\alpha_0/2}, \quad (8)$$

where  $C$  stands for the critical value matrix,  $S_0$  is the initial standard deviation value,  $Q_{vv}$  consists of the variance-covariance factors of the residuals,  $P$  is the weighted observations matrix,  $f$  is the degree of freedom and  $\alpha_0$  is the reference level value. It is obvious that the critical value for each observation is calculated by computing the average of the computed critical values mentioned in Eq. (8).

## 1.2. Fuzzy approach

Recently, the outlier detection approach using Fuzzy method has been employed by (Sisman *et al.* 2012) for 3D coordinate transformation. The overall procedure in Fuzzy methodology is similar to the complementary and improved operators in Fuzzy logic algorithm which we have used in this study. To manipulate this approach on point clouds, we considered the Fuzzy methodology in a distributed based approach domain. In fact, the essential idea in Fuzzy logic algorithm which was proposed firstly by (Zadeh 1965) can broaden the classical counting boundaries that would use the suitable subscription functions and then link them to the corresponding variables. In the classical set of numbers, the subscribed value for each element was 1 if that was available in the corresponding set, otherwise it must be considered as 0. Meanwhile, the element value in this algorithm indicates the element's dependency degree with respect to the entire set's elements belongs to the span of 0 to 1 written as (0, 1).

In this paper, the Fuzzy technique is used to detect the outliers based on the subscription function values by taking advantage of the residuals' features. Since the real errors are from the unknown observations in outlier detection process, the residuals and redundancy values are used here as the testing tools.

Accordingly, if we suppose that  $\Delta$  is the vector of outlier in a linear functional model, the mathematical connections between residuals vector and the vector of errors in the observations can be written as follows:

$$V = -(I - A \cdot Q_X \cdot A^T \cdot Q_L^{-1}) \Delta = -Q_V \cdot Q_L^{-1} \Delta = -R \Delta, \quad (9)$$

where  $R$  denotes the redundancy matrix which determines the relation between errors of observations and the residuals. Also, Eq. (9) in the form of matrix can be rewritten as:

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = - \begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \\ r_{21} & r_{22} & r_{23} & \cdots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & r_{n3} & \cdots & r_{nn} \end{bmatrix} \times \begin{bmatrix} \Delta_1 \\ \Delta_2 \\ \Delta_3 \\ \vdots \\ \Delta_n \end{bmatrix}. \quad (10)$$

In Eq. (10), it is clear that the residuals are affected by those existed total errors in the observed parameters. More importantly, given Eq. (9) and Eq. (10), if the trace of matrix  $R$  may be available (for uncorrelated observations and being a real value), it would give us the degree of freedom for those total parameters. Although the  $R$  matrix is not invertible and thus it is a big problem to apply the inverse matrix transformation on both sides of Eq. (10).

Here the test values for all residuals are computed based on Eq. (9) and Eq. (10) during the first round of LSS. On the other hand, these test values are figured out from a range of iterative procedures in the classical methods of outlier detection. Based on this, those residuals larger than the critical test values are being considered as the outliers in the next round of LSS adjustment. Therefore, according to the Fuzzy approach, the residuals can be divided in two general groups, after comparison of them with the following test values:

- Observations with normal residuals and test values less than the critical statistical values; these are defined here as  $N(v_i)$ ;
- Observations with abnormal residuals which their test values are larger than the critical statistical values; these are defined as  $M(v_i)$ .

According to the general statistical viewpoints, those observations with test values even a little bit greater than the assumed criteria belong to the abnormal observations. This can cause an uncertainty deficiency which will be resolved eventually by using subscribed value concept. In fact, in Fuzzy methodology, the subscribed functions for those residuals lower than the critical limitation are evaluated lonely. Besides, those values for subscribed functions are considered 0 for each member of  $N$ . Similarly, the values for each member of  $M$  with test value greater than the critical limitation will get value from the span of (0,1). Therefore, the subscribed functions related to the residuals are defined as:

$$m_{\tilde{M}}(v_i) = \begin{cases} 0; & W_i \leq N_{1-\alpha/2}(0, 1) \\ \frac{1}{1 + r_{ii} \left( \frac{\alpha}{W_i - N_{1-\alpha/2}(0, 1)} \right)^2}; & W_i > N_{1-\alpha/2}(0, 1) \end{cases}, \quad (11)$$

where those subscribed function values for the related observations affected by the outliers can be given as follows:

$$m_{\tilde{N}}(v_i) = 1 - m_{\tilde{M}}(v_i). \quad (12)$$

In order to discriminate the Fuzzy subscribed relations within the entire observations, we have used the redundancy matrix where its elements have been normalized as Eq. (13) as follows:

$$\tilde{r}_{ij} = \frac{|r_{ij}|}{\max |r_{ij}|}; \quad i, j = 1, 2, 3, \dots, n, \quad (13)$$

where  $r_{ij}$  denotes the normalized value for each element of matrix  $R$  (Eqs. 9 and 10).

Therefore, by normalizing the entire elements of redundancy matrix, the relative redundancy matrix will give values between 0 and 1. Moreover, the relative matrix's rows and columns indicate the relatively effects of observed errors on the residuals which would be caused through the influence of every residual on each observation's error. Given this condition, it is possible to divide the observed errors as two separated groups as matrix  $A$  which includes the observed errors with the most effectiveness on the abnormal residuals; and matrix  $B$  which includes the observed errors with least effectiveness on the normalized residuals. Thus, in order to get the subscribed values for matrixes of  $A$  and  $B$ , the relative redundancy matrix defined as  $\tilde{R}$  matrix and the subscription functions of  $m_N$  and  $m_M$  are being used.

Additionally, it is supposed that the subscribed values of  $A$  and  $B$  matrixes are  $m_A(\Delta_i)$  and  $m_B(\Delta_i)$ , respectively, then the corresponding subscribed function values have been calculated as the following steps:

If we consider  $\alpha$  in a Fuzzy set of  $M$ , the maximum relative effectiveness of observed error of  $i$ -th on the residuals will be as  $m_M(v) \leq 0.5$ , then:

$$\tilde{r}_{mi} = \max(|\tilde{r}_{ki}|); \quad v_k \in M_{0.5}. \quad (14)$$

Therefore, the subscription values for observed errors can be computed as:

$$m_{\tilde{A}}(\Delta_i) = \tilde{r}_{mi} \cdot m_{\tilde{M}}(v_i). \quad (15)$$

Similarly, the subscription values for  $B$  can be written as:

$$m_{\tilde{B}}(\Delta_i) = 1 - \tilde{r}_{ni} \cdot m_{\tilde{N}}(v_i), \quad (16)$$

where  $r$  stands for the maximum relative value for the observed error of  $i$ -th with taking into the consideration the following condition:

$$m_{\tilde{N}}(v) \geq 0.5 \quad \text{and then} \quad \tilde{r}_{ni} = \max(|\tilde{r}_{ki}|); \quad v_k \in N_{0.5}. \quad (17)$$

In Eq. (17), those observations with the greatest effectiveness are those which whether hold the maximum effect on abnormal residuals or hold the

minimum effect on the normal residuals or even both of the aforementioned conditions.

However, the percentage of being outlier for observation  $L_i$  is the maximum value of subscribed values of  $m_A(\Delta_i)$ , but according to the Fuzzy theory, the combined set of two sets of  $A$  and  $B$  creates a new set which can be defined by  $H$  set as the following equation:

$$m_{\tilde{H}}(\Delta_i) = \max(m_{\tilde{A}}(\Delta_i), m_{\tilde{B}}(\Delta_i)). \quad (18)$$

Regarding to the subscribed values of  $H$ , it is clear that the percentage of an outlier is increasing whenever the corresponding value in  $H$  matrix increases. Thus, it will be up to the decision criteria which this observation can be considered as an outlier, a normal observation and/or a subscribed value.

Moreover, in order to determine a clear boundary, here the defuzzification limitation is defined as:

$$C_H = \frac{\sum P_i m_{\tilde{H}}(\Delta_i)}{\sum P_i}, \quad (19)$$

where in Eq. (19):

$$P_i = \begin{cases} \tilde{r}_{mi}; & m_{\tilde{H}}(\Delta_i) = m_{\tilde{A}}(\Delta_i) \\ \frac{1}{m_{\tilde{N}}(v_i)} - \tilde{r}_{ni}; & m_{\tilde{H}}(\Delta_i) = m_{\tilde{B}}(\Delta_i) \end{cases}. \quad (20)$$

Therefore, if we consider a subscribed value of  $m_{\tilde{H}}(\Delta_i)$  with defuzzification limitation value of  $C_H$ , an observation which verifies the following equation will be considered as an outlier:

$$m_{\tilde{H}}(v) \geq C_H. \quad (21)$$

All in all, by using Eq. (21) and comparison with  $m_{\tilde{H}}(v)$  for each critical observation as  $C_H$  will provide us with sufficient information to make a true decision about considering an observation as an outlier or a normal one.

According to above definitions, the proposed methodology employed in this study and the corresponding steps are presented in Figure 1 as a flowchart.

### 1.3. Data and observations: ground-based laser scanner and randomly simulated distributed point cloud

In order to evaluate the mentioned Baarda method and Fuzzy approach practically, two groups of point cloud including ground-based laser scanner data and randomly 3D simulated points are used.

The used ground-based laser scanner data in this study has been provided by a laser scanner

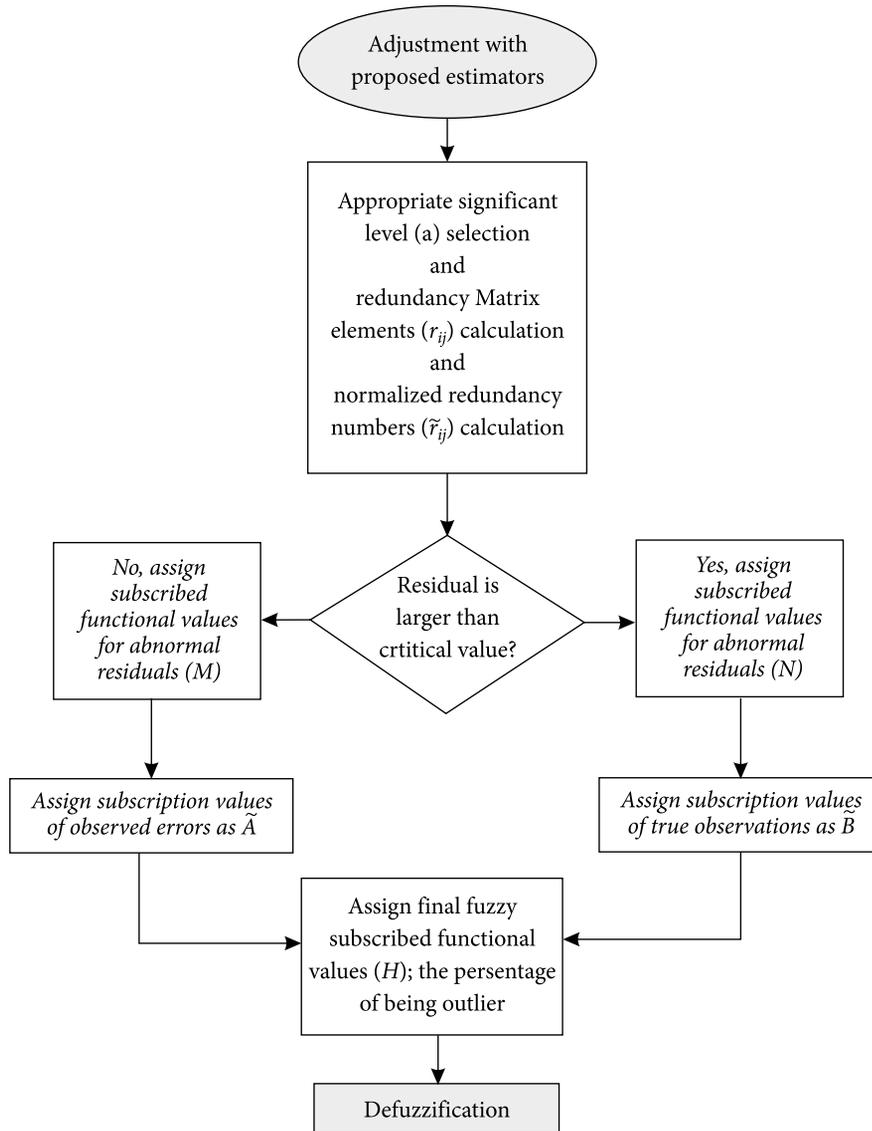


Fig. 1. Flowchart of proposed Fuzzy methodology in this study

manufactured by RIEGL Company (RIEGL VZ-4000). This laser scanner has been employed to collect 3D points from a side wall – ScanPos – of a construction in Ghaleh-Porteghali region (Portuguese Castle: Latitude: 26.78° N, Longitude: 54.89° E, located in Hormozgan county, Iran). Next, the collected initial observations have been processed by RiSCAN-PRO software and then extracted in an optimized extension which can be applied into further programming frame and statistical data processing.

The randomly simulated point clouds have been prepared according to the following steps; firstly a range of random  $x$  and  $y$  coordinates are being considered in a mesh grid frame with an equal distance with respect to each other. In spite of ground-based laser scanner data – which is not an easy task to observe the pre-established control coordinates in an exact mesh

frame – we initially supposed that  $x$  and  $y$  coordinates are located in the same mesh frame. Furthermore,  $z$  coordinates for these simulated point clouds are also the random values according to NDF features. Besides, here the fitting method and the random errors are based on LSS and NDF features, respectively. Therefore, there are no external effects that can be added to the  $z$  values systematically which subsequently will not affect those justification assumptions existed in LSS method.

## 2. Results and discussions

In this study, the Baarda method and Fuzzy approach are utilized for point cloud's robust outlier detection. The statistical reliability of these approaches is also evaluated numerically. For such evaluation, we have employed these two methods on two data groups

including the point cloud derived from the ground-based laser scanner field experiment and randomly simulated data sets. In the following two sections, we present the obtained results and analysis for each data group in details.

### 2.1. Baarda method and Fuzzy approach utilization

#### 2.1.1. Ground-based laser scanner point cloud

As mentioned above, Baarda method works based on computing the distance of each point from the whole point clouds. Although it would consider a known

realized reference level (for instance,  $\alpha = 0.1$ ), but the points with the lengths more than the critical value must be classified as the outliers. On the other hand, the Fuzzy approach works based on the outputs of conventional outlier detection methods (mentioned above). It would work only those observations which assumed to be outliers in that conventional outlier detection method.

According to this, Figures 2 and 3 present the Baarda method and Fuzzy approach applied to the point cloud of ground-based laser scanner field experiment, respectively.

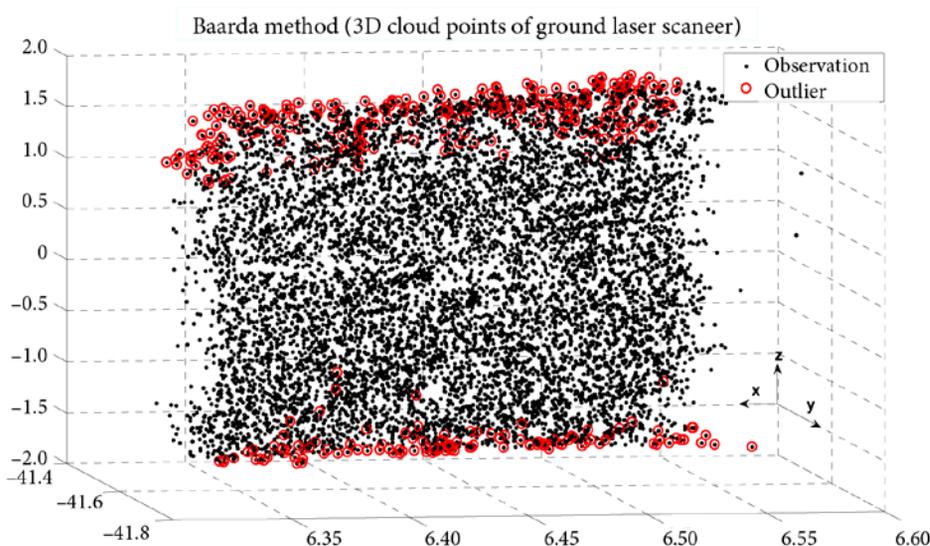


Fig. 2. Outliers detection based on Baarda method applied into the 3D point clouds of the ground-based laser scanner ( $x$ ,  $y$  and  $z$  [m] are assumed in a local coordinate system).

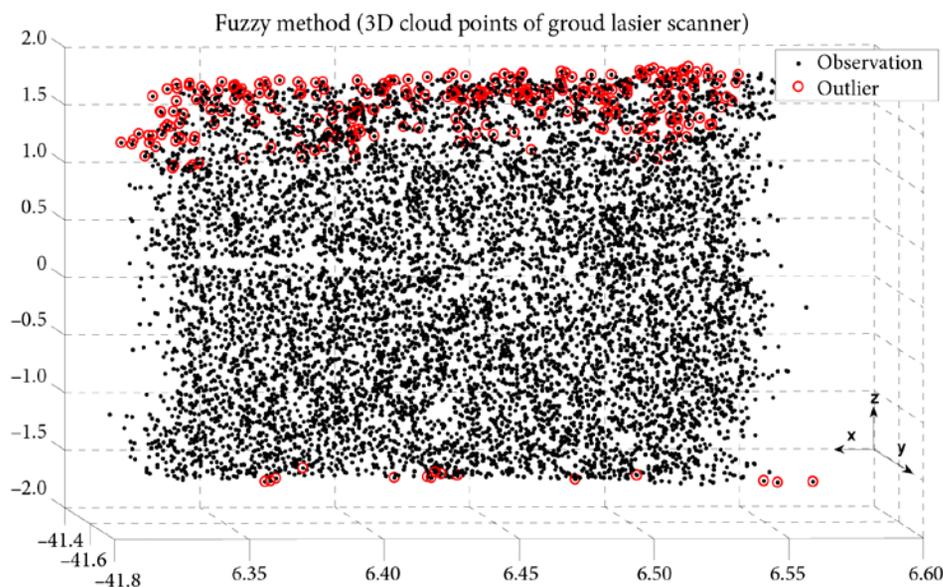


Fig. 3. Outliers detection based on Fuzzy approach applied into the 3D point clouds of the ground-based laser scanner ( $x$ ,  $y$  and  $z$  [m] are assumed in a local coordinate system).

### 2.1.2. Randomly simulated point cloud

In order to apply two aforementioned approaches for randomly simulated point cloud, first a 3D surface with an optimized polynomial formula should be fitted to each point cloud data set. In fact, the goodness of fit for this polynomial can be evaluated from two different approaches; the optimized degree of polynomial and the corresponding justification's type which has been used in the fitting procedure. According to this, by considering the type and the characteristics of each data set, there will be an optimized degree for the employed polynomial. Thus, as much as the selected polynomial's degree is considered close to an optimized value, the fitting procedure will be more accurate as well. In this study we have used the LSS method to fit the suitable polynomial to each data set, since the LSS is one of the best solutions in terms of statistic properties. To do so, we firstly calculated the entire polynomial coefficients using the whole data set points and then these coefficients are being improved in each step especially by removing the existing outliers.

Accordingly, we initially calculated the coefficients of fitted polynomial for each point cloud dataset. Next, the distance between each point and the optimal polynomial surface is being computed numerically. Then, we applied separately the Baarda and Fuzzy method as showed in Figures 2 and 3, respectively. Based on the type of simulated point cloud which is a mesh grid, the used polynomial in the fitting procedure is a polynomial with degree 1 and its equation consists of three coefficients as follows:

$$z = ax + by + c, \quad (22)$$

where  $a$ ,  $b$ ,  $c$  are the constant coefficients and  $x$ ,  $y$  and  $z$  are the coordinates of 3D points in each data set. Since the used point cloud has a large size, the probability of existing outliers is very high. It is therefore necessary to determine an optimal polynomial surface in each step of computation especially after eliminating the outliers. This optimal surface determination's procedure has been made continuously until there will be no outlier in the data sets.

Next, we applied Baarda method and Fuzzy approach into random sets of simulated could point. Figure 4 presents the outputs of Baarda methodology for the simulated point cloud data sets. Similarly, Figure 5 represents the detected outliers after applying the Fuzzy approach.

Considering the results of these procedures, in comparisons and discussions section, we discuss the statistical data analysis, accuracy and sensitivity of these two approaches on the next section in details.

### 3. Comparisons and discussions

Since it is important to apply the Baarda method and Fuzzy approach in different levels of point cloud size, we have considered three sets of point clouds as set A, B and C for both ground-based laser scanner as well as random simulated point cloud. According to this, Table 2 represents the comparisons of Baarda and Fuzzy methodologies in detecting outliers for set A, B and C derived from the ground-based laser scanner field experiment point cloud.

As it can be understood from Table 2, Fuzzy method performs more precise to detect the outliers.

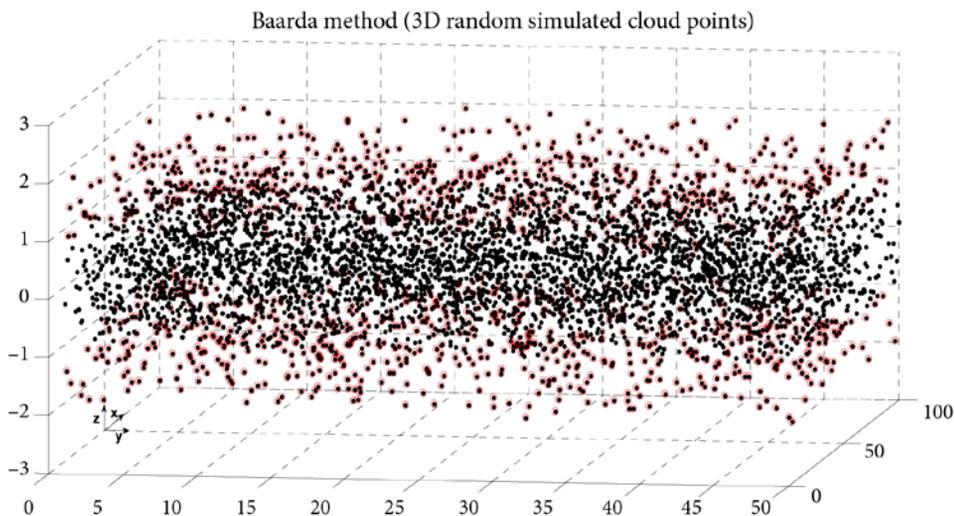


Fig. 4. Outliers detection based on Baarda method applied into the randomly simulated 3D point clouds ( $x$ ,  $y$  and  $z$  [m]).

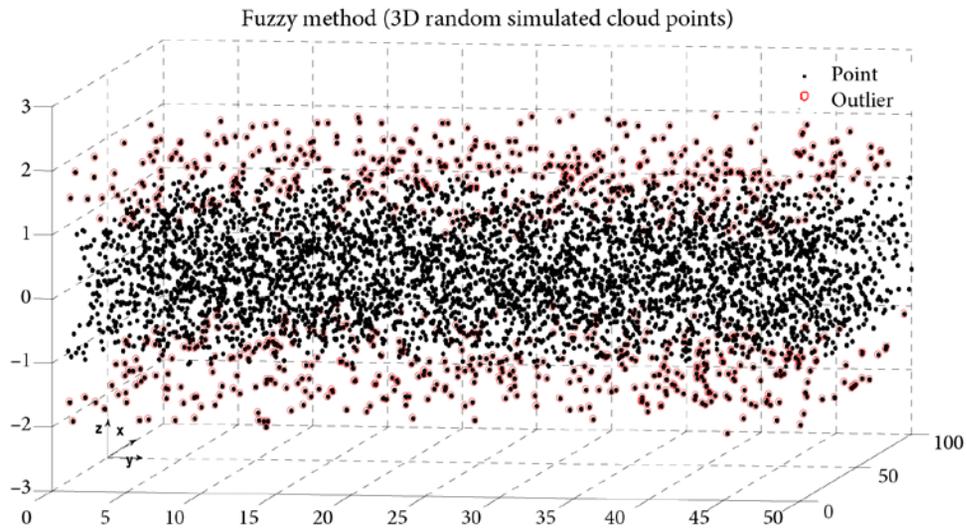


Fig. 5. Outliers detection based on Fuzzy approach applied into the randomly simulated 3D point clouds (x, y and z [m]).

Table 2. Comparison of Baarda method and Fuzzy approach to detect outliers for 3D laser scanner point clouds

Methodology	Number of point clouds	Point Clouds		
		Set A	Set B	Set C
		1115	4565	8861
Baarda	Number of detected outliers	8	164	414
	Estimated variance factor	0.57472	0.84044	0.88872
Fuzzy	Number of detected outliers	8	113	313
	Estimated variance factor	0.57472	0.85547	0.90455

Table 3. Comparison of Baarda method and Fuzzy approach for using LSS to calculate the unknown coefficients in surface fitting procedure

		Methodology		
		X (without outlier detection)	X (Baarda)	X (Fuzzy)
set A	a	7.1722	6.8737	6.8737
	b	15.8404	16.7952	16.7952
	c	616.1665	657.8845	657.8845
set B	a	14.2431	16.709	18.1179
	b	-6.5827	-6.5581	-7.4604
	c	-364.9788	-379.7518	-426.4424
set C	a	17.1877	24.4088	23.2743
	b	-17.6444	-24.8159	-23.6652
	c	-845.0367	-1190.138	-1135.044

In fact, the Fuzzy method keeps those points with relevant estimators located exactly near the critical values. Although randomly simulated point cloud has this possibility to insert a certain number of outliers into the points, but it is feasible to measure the successfulness of the applied methods based on this priori information. Conversely, the number of real outliers in the ground-based laser scanner field experiment point clouds is unknown. In order to deal with this, the reliability of each method will be determined through analyzing the secondary factor variance ( $\hat{\sigma}_0^2$ ) as well as measuring the rate of changes in the unknowns ( $\hat{a}, \hat{b}, \hat{c}$ ) (in initial step and also after removing the outliers). Accordingly, by applying the Fuzzy method and removing the detected outliers, the secondary factor variance is approaching to 1, gradually. Besides, it can be seen clearly from the randomly simulated datasets given in Table 3 that the changes occurred on the estimated unknowns is considerable enough.

In fact, these changes are close to each other after removing the outliers by using whether Baarda method or Fuzzy approach. On the other hand, the changes between these estimated unknowns are significant before and after removing the outliers using whether Baarda or Fuzzy method. These great changes imply that not only the outlier detection process can impair the estimated results, but also different outlier detection procedures have considerably different results comparing to each other.

Table 4 gives the comparison of two methods applied into the randomly simulated 3D point cloud data sets.

Table 4. Comparison of Baarda method and Fuzzy approach to detect outliers for randomly simulated 3D point clouds

	Methodology	Number of Simulated Point clouds		
		Set A	Set B	Set C
		500	5000	10000
Estimated Factor Variance	Before outlier detection	1.1530	0.5649	0.8521
	Baarda	1.0150	0.6644	0.9068
	Fuzzy	1.0079	0.7254	0.9343

Since it is possible to analyze these point cloud data sets based on the numbers and type of outliers which are already known, the results can be evaluated from many aspects. First, as we see that if the number of outliers is very large (less than 20 percent of the data set), both Fuzzy and Baarda method are going to provide successful results. However, if the number of observations may increase in particular more than 20 percent of the data set's size, the Baarda method in a predetermined critical value (referred as  $\alpha$ ) would like to remove the true observations when it parallelly eliminates the outliers. Therefore, it practically causes a reduction in the number of correct observations which can be seen clearly among all three data sets (set A, B and C). In addition to this, if we suppose that the number of inserted outliers is constant, by considering the predetermined critical value less than 90 percent ( $\alpha \geq 0.1$ ), the Baarda method will remove both the possible outliers besides a percentage of correct observations. This characteristic of Baarda method can decrease the accuracy of the adjustment process, estimating steps and the fitting procedure. Consequently, it can be understood that the Fuzzy method whether in the normal condition or in the critical limitation one, would work more reasonable and accurate. Also, the estimated values using Fuzzy method are more similar to the real ones with less reduction in the number of true observations (Table 5).

According to Table 2 and Table 3, the number of detected outliers which are found by Baarda method is equal to or greater than Fuzzy method in both data group. This demonstrates that Fuzzy method will be able to detect those outliers located far enough from the critical value. Moreover, in low size point cloud (e.g. set A and A), Baarda method behaves very similar to the Fuzzy method; while the performance of Fuzzy method for set A and A seems to be more reasonable in both data group. According to the selected

Table 5. Comparison of Baarda method and Fuzzy approach to detect inserted outliers to the randomly simulated 3D point clouds based on critical limitation value variability

	$\alpha$	Inserted Outlier	Detected Outliers	
			Baarda	Fuzzy
set A (500)	0.15	10	10	10
		40	40	40
		60	60	60
		80	80	80
		100	140	100
		120	320	120
		160	500 (all points)	160
set B (5000)	0.15	100	100	100
		400	400	400
		800	800	800
		900	900	900
		1000	1300	1000
		1200	3100	1200
set C (10000)	0.15	100	100	100
		1000	1000	1000
		2000	2000	2000
		2200	2200	2200
		2300	2420	2300
		2450	7510	2450
		2500	7900	2500
	Inserted Outlier	$\alpha$	Detected Outliers	
			Baarda	Fuzzy
set A (500)	100	0.01	100	100
		0.05	100	100
		0.1	100	100
		0.15	140	100
		0.2	240	100
		0.25	334	100
set B (5000)	1000	0.01	1000	1000
		0.05	1000	1000
		0.1	1000	1000
		0.15	1300	1000
		0.2	2400	1000
		0.25	3300	1000
set C (10000)	2000	0.01	2000	2000
		0.05	2000	2000
		0.1	2000	2000
		0.15	2000	2000
		0.2	3700	2000
		0.25	7300	2000

critical value and secondary factor variance ( $\hat{\delta}_0^2$ ), this comparison presents that the Fuzzy method in larger point cloud's size (e.g. B, B, C and C) approaches to the acceptable value of 1 as it improves the adjustment computational procedures and modeling parallelly.

## Conclusions

This paper presents an up-to-date application and detailed comparison of Baarda method and Fuzzy approach as outlier detection methodologies for 3D point clouds derived from ground-based laser scanner field experiment and randomly simulated point cloud data sets. Since there is a range of problems during the data acquisition by ground-based laser scanner, it is not easy to re-establish the data acquisition procedure in order to detect the possible outliers especially it is highly costly. Thus, it is necessary to focus on data processing methodologies to conclude a more reasonable and applicable approach. In this study, the results showed that the outputs of these two methods are similar to each other while to decide if the critical value is an outlier or a normal observation (considering corresponding estimators), they are completely different. In fact, the Fuzzy method detects the outliers perfectly, but it is not easy to confirm whether it can detect all the possible outliers within the entire observations or not. Besides, the outlier's classification can be done using the subscribed function value of each observation. Moreover, the uncertainty of Fuzzy method is less than those conventional statistical methods in particular the Baarda method (as mentioned in the Table 2, Table 3, Table 4 and Table 5). It should be noted here that the output of each method may vary due to many convoluted parameters such as the size and density of point cloud and the percentage of holding the possible outliers. Also, since the optimal fitted model is computed here through LSS, a single outlier will effect on the whole model formation. Therefore, detecting such an outlier is playing a very important role, but the certainty of statistical tests to detect outliers is not acceptable. On the other hand, the general formation of solution's model is constant during the applying of Fuzzy method. As it mentioned, the optimal fitted model is changing for the common outlier detection methods during each step of LSS. The latter statement can be considered as a technical and novel characteristic of Fuzzy method for the point cloud's outlier detection concept.

## References

- Aliosmanoğlu, Ş.; Akyılmaz, O. 2002. *A comparison between statistical & fuzzy techniques in outlier detection: Vistas for geodesy in the new millennium*. Springer.  
[http://dx.doi.org/10.1007/978-3-662-04709-5\\_63](http://dx.doi.org/10.1007/978-3-662-04709-5_63)
- Baarda, W. 1968. A testing procedure for use in geodetic networks, *Rijkscommissie voor Geodesie* 1: 5–97.
- Baarda, W.; Commission, N. G. 1967. Statistical concepts in geodesy, *Rijkscommissie voor Geodesie* 2(4): 5–74.
- Berberan, A. 1995. Multiple outlier detection. A real case study, *Survey review* 33: 41–49.
- Breunig, M. M.; Kriegel, H. P.; Ng, R. T.; Sander, J. 2000. LOF: identifying density-based local outliers, *Association for Computing Machinery Sigmod Record* 29(2): 93–104.  
<http://dx.doi.org/10.1145/342009.335388>
- Gökalp, E.; Güngör, O.; Boz, Y. 2008. Evaluation of different outlier detection methods for GPS networks, *Sensors* 8: 7344–7358. <http://dx.doi.org/10.3390/s8117344>
- Hampel, F. R.; Ronchetti, E. M.; Rousseeuw, P. J.; Stahel, W. A. 2011. *Robust statistics: the approach based on influence functions*. John Wiley & Sons.
- Huber, P. J. 2011. *Robust statistics*. Springer.  
[http://dx.doi.org/10.1007/978-3-642-04898-2\\_594](http://dx.doi.org/10.1007/978-3-642-04898-2_594)
- Jahandideh, S.; Azizi, A.; Najibi, N. 2014. Numerical evaluation and application-oriented analysis for forward and inverse rational function models of terrain-independent case in satellite imagery, *Geodesy and Cartography* 40: 99–109.  
<http://dx.doi.org/10.3846/20296991.2014.962731>
- Johnson, T.; Kwok, I.; Ng, R. 1998. Computation of 2-dimensional depth contours, in *Proceedings of 4<sup>th</sup> International Conference on Knowledge Discovery and Data Mining*, 27–31 August 1998, New York, USA, 224–228.
- Koch, K. R. 1999. *Parameter estimation and hypothesis testing in linear models*. Springer.  
<http://dx.doi.org/10.1007/978-3-662-03976-2>
- Monhor, D.; Takemoto, S. 2005. Understanding the concept of outlier and its relevance to the assessment of data quality: Probabilistic background theory, *Earth, Planets, and Space* 57: 1009–1018. <http://dx.doi.org/10.1186/BF03351881>
- Muñoz-García, J.; Moreno-Rebollo, J.; Pascual-Acosta, A. 1990. Outliers: a formal approach, *International Statistical Review* 58(3): 215–226. <http://dx.doi.org/10.2307/1403805>
- Najibi, N.; & Arabsheibani, R. 2013. Snow-covered surface variability and DEM generation using aerial photogrammetry in Mount Odin, Canada, *Geodesy and Cartography* 39: 113–120. <http://dx.doi.org/10.3846/20296991.2013.823704>
- Najibi, N.; & Jin, S. 2013. Physical reflectivity and polarization characteristics for snow and ice-covered surfaces interacting with GPS signals, *Remote Sensing* 5: 4006–4030.  
<http://dx.doi.org/10.3390/rs5084006>
- Papadimitriou, S.; Kitagawa, H.; Gibbons, P. B.; Faloutsos, C. 2003. LOCI: Fast outlier detection using the local correlation integral, in *Proceedings of 19th International Conference on Data Engineering*, 5–8 March 2003, Bangalore, India, IEEE, 315–326. <http://dx.doi.org/10.1109/icde.2003.1260802>
- Rao, C. R.; Toutenburg, H. 1995. *Linear models*. Springer.  
<http://dx.doi.org/10.1007/978-1-4899-0024-1>
- Sisman, Y.; Dilaver, A.; Bektas, S. 2012. Outlier detection in 3D coordinate transformation with fuzzy logic, *Acta Montanistica Slovaca* 17(1): 1–8.

- Sotoodeh, S. 2006. Outlier detection in laser scanner point clouds. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36: 297–302.
- Sun, W. 1994. A new method for localisation of gross errors, *Survey Review* 32: 344–358.  
<http://dx.doi.org/10.1179/sre.1994.32.252.344>
- Vaniček, P.; Wells, D. 1972. *The least squares approximation and related topics*. Fredericton: University of New Brunswick.
- Zadeh, L. A. 1965. Fuzzy sets, *Information and control* 8: 338–353. [http://dx.doi.org/10.1016/S0019-9958\(65\)90241-X](http://dx.doi.org/10.1016/S0019-9958(65)90241-X)

---

**Reza ARABSHEIBANI**. M.Sc. in Geospatial Information System (GIS), School of Surveying and Geomatics Engineering, College of Engineering, University of Tehran, North Kargar Ave., P.O. Box: 11155-4563, Tehran, Iran. Phone +9821 66493046 & +98 917 951 9274. E-mail: [rasheibani@ut.ac.ir](mailto:rasheibani@ut.ac.ir). B.Sc. in Surveying and Geomatics Engineering, School of Surveying and Geomatics Engineering, Faculty of Engineering, University of Tehran.

Research interests: geospatial information system, semantic web, outlier detection, optimization, tide and tidal currents analysis, volunteered geographic information (VGI), spatial modelling.

---

**Abbas ABEDINI**. Assistant Professor, School of Surveying and Geomatics Engineering, College of Engineering, University of Tehran, Iran. Phone: +989123473735. E-mail: [aabedeni@ut.ac.ir](mailto:aabedeni@ut.ac.ir). M.Sc, Photogrammetry and Geoinformatic Engineering, Stuttgart University of Applied Sciences. B.Sc. in School of Surveying and Geomatics Engineering, College of Engineering, University of Tehran, Tehran, Iran. Research interests: geomatics engineering, hydrology engineering, designing of route.

---

**Yousef KANANI SADAT**. PhD student, School of Surveying and Geomatics Engineering, College of Engineering, University of Tehran, Iran. E-mail: [yousefkanani@ut.ac.ir](mailto:yousefkanani@ut.ac.ir). M.Sc. in GIS, School of Surveying and Geomatics Engineering, College of Engineering, University of Tehran, Iran. B.Sc. in Surveying and Geomatics Engineering, College of Engineering, University of Tehran, Tehran, Iran. Research interests: spatial modelling, spatial data mining, volunteered geographic information (VGI), spatial cognition.