Advanced Geomatics– 2023; 3(1); 28-32

Advanced Geomatics

http://publish.mersin.edu.tr/index.php/geomatics/index

e-ISSN: 2791-8637

The Outlier Detection with Robust Methods: Least Absolute Value and Least Trimmed Square

Ulku Kırıcı Yıldırım*[1], Hasan Dilmaç [1], Yasemin Şişman [1]

[1]*Ondokuz Mayis University, Engineering Faculty, Department of Geomatics Engineering, Samsun, Tükiye*

Keywords

Adjustment,
Outlier Measurement Test,
LS Method,
LAV Method,
LTS Method.

Research Article

Received : 11.09.2022
Revised : 16.02.2023
Accepted : 09.03.2023
Published : 31.03.2023

Abstract

In geodesy and surveying, measurements usually have errors. These errors are called outlier measurements. In order to determine these points, outlier measurement test is performed. There are many different methods used to determine outlier measurement. The least squares (LS) method is the most common method to estimate the unknowns from outlier measurements. However, LS method can be easily affected by outliers which may cause wrong results. Classical outlier tests and robust methods are the two main approaches to detect outliers or reduce their effect. There are a lot of robust methods in literature. In this study, least square method (LS), least absolute value (LAV) and the least trimmed squares (LTS) are discussed. To compare the outlier performances of the methods, real data points are used to create a surface with a 2nd degree polynomial.

1. Introduction

Adjustment is important in surveying because the measurements which are made to determine the unknowns usually have errors. Therefore, redundant measurements are needed to increase the precision of the computed unknowns (Ghilani, 2017). Among the many methods, the least squares (LS) has been generally used to adjust the Gauss-Markov Model (GMM) in surveying, geodesy and different fields (Fang, 2015). However, outliers could happen in the measurements due to different reasons and they may affect the results, thus causing wrong assumptions (Erdogan, 2014). So, outliers must be detected or their effect must be reduced by using some methods. Classical outlier tests (Baarda, 1968; Pope, 1976; Koch, 1999) and robust methods are two main approaches in geodesy to detect outliers or reduce their effect (Sisman, 2010).

Classical outlier tests could be ineffective if outlier number is large. In this case, outliers can remain undetected. Moreover, even correct measurements can be detected as outlier wrongly (Berné Valero & Baselga Moreno, 2005). At this point, robust methods (Huber, 1981; Hampel et al. 1986; Rousseeuw, 1984; Rousseeuw & Leroy, 1987) are developed to be insensitive to outliers. Also, detection of outliers can be done by looking at the residuals from a robust method (Hekimoglu & Erenoglu, 2009). There are many robust methods such as M-estimators, least absolute value (LAV) and Generalized M-estimators, least median of squares (LMS) and the least trimmed squares (LTS) and so on (Hekimoğlu, 2005).

In this study, outlier measurements were determined by using a data set from Ondokuz Mayis University in Samsun. The solution was made with LS, LAV and LTS methods. Results were compared.

2. Method

The outlier performances of LS, LAV and LTS methods are analyzed in the study. They are briefly introduced in three different headings.

* **Corresponding Author**

*(ulku.kirici@omu.edu.tr) ORCID ID 0000-0002-3569-4482
(hasan.dilmac@omu.edu.tr) ORCID ID 0000-0001-6877-8730
(ysisman@omu.edu.tr) ORCID ID 0000-0002-6600-0623

Cite this;

Yıldırım, U. K., Dilmaç, H. & Şişman, Y. (2023). The Outlier Detection with Robust Methods: Least Absolute Value and Least Trimmed Square. Advanced Geomatics, 3(1), 28-32.

2.1. The Least Square Method (LS) and Outlier Detection Procedure

The least squares method is explained by Carl Friedrich Gauss in 1795 and Legendre in 1805. LS objective function is $\|Pvv\| = [Pvv] = \min$. Unknown parameters are calculated with the following equation in this method (Sisman, 2014).

$$\underline{X} = \left(\underline{A}^T P \underline{A}\right)^{-1} \underline{A}^T P \underline{\ell} \quad (1)$$

Root mean square error (RMSE);

$$m_0 = \pm\sqrt{\frac{V^T PV}{f}}\ ;\ f = n - u \quad (2)$$

The measurement errors of the LS method influence the residual of other calculations. Thus, this correction value may not always be due to an error in the measurement. This situation is called the spread and storage effect of LS method. Different solution methods can be conducted for the analysis of spread and storage method (Ayan,1992).

The point with the highest V value is removed from the cluster. The outlier measurement test is repeated with the remaining points. This process is continued iteratively until there is no outlier measurement (Kirici, Sisman, 2015).

2.2. The Least Absolute Value Method (LAV) and Outlier Detection Procedure

In the classical Gauss-Markov model, the unknown parameter (x) for a linear (linearized) parametric adjustment is determined based on the following functional and stochastic models (Simkooei A.A., 2003).

$$\begin{aligned} l + v &= A.x \\ D^T.x &= 0 \\ P &= Q_l^{-1} = \sigma_0^2 C_l^{-1} \end{aligned} \quad (3)$$

$v_{n\times1}$:vector of residual, $l_{n\times1}$: vector of observation, $A_{n\times u}$: rank deficient design matrix, $P_{n\times n}$: weight matrix of observations, $D_{u\times d}$: datum matrix of the network added to complete the rank deficiency of the design matrix, $O_{d\times1}$: zero vector, $C_{j(n\times n)}$: covariance matrix of observations, $Q_{l(n\times n)}$: cofactor matrix, σ_0^2: priori variance factor

LAV objective function is $||Pv|| = [P|v|] = min$. It contains the L1 Norm method and unknown parameters such as $\underline{X}$ and $\underline{V}$. The new unknowns for linear programming are arranged as follows.

$$\begin{aligned} X &= X^+ - X^- \ ; \quad X^+, X^- \geq 0\ , \\ V &= V^+ - V^- \ ; \quad V^+, V^- \geq 0 \end{aligned} \quad (4)$$

The mathematical model and constraint equation for linear programming in the solution according to the LAV objective function are as follows (Bektas ve Sisman 2010).

$$\begin{bmatrix} A & -A & -I & I \end{bmatrix} \begin{bmatrix} X^+ \\ X^- \\ V^+ \\ V^- \end{bmatrix} = [\ell]\ , \qquad (5)$$
$$f = b^T X = [P|V|] = P^T V = p^T [V^+ \quad V^-] = min.$$

By looking at the calculated V values at the end of the process, outlier measurement can be easily determined (Kirici 2016).

2.3. The Least Trimmed Square (LTS) and Outlier Detection Procedure

The LTS method (Rousseeuw & Leroy, 1987) is a high breakdown estimator. The objective function of the LTS is given as:

$$Min \sum_{i=1}^{h} P_i v_i^2 \quad (6)$$

Here, h, P *and* v represent the trimming parameter, the weight of the measurements and the residuals of the measurements respectively. When $h = n$ (measurement number), this is same as LS method. h is usually set to a constant number smaller than n (Mount et al., 2014). For the best robustness $h = [n + u + 1]/2$ (Rousseeuw & Leroy, 1987).

In this method, the h measurements which have the smallest squared residuals are searched for out of n measurements (Dilmac and Sisman, 2023). When the number of possible $\binom{n}{h}$ subsets is relatively large, a full searching of all possible subsets is prohibitive. Several algorithms (Hawkins, 1994; Atkinson & Cheng, 1999; Li, 2005; Koch et al., 2017) are proposed to overcome this issue. For this study, the FAST-LTS algorithm (Rousseeuw & Driessen, 2006) is discussed. A workflow that briefly explains this algorithm is given below (Figure 1).

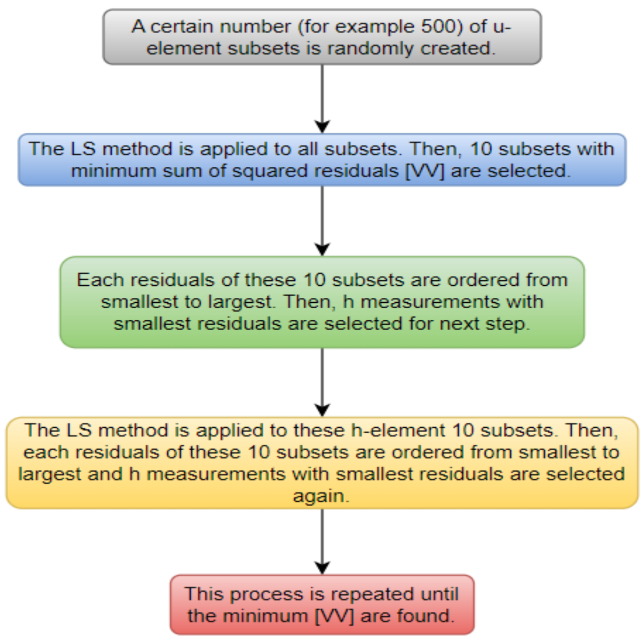

Figure1. The workflow of FAST-LTS algorithm.

First, least squares is applied with 500 randomly generated sets with u elements. From these 500 subsets, 10 subsets with the smallest [VV] are selected. The residuals for all measurements are calculated using the estimated X of these 10 subsets. Then, the residuals of these 10 subsets are ordered from the smallest to the largest and h measurements with smallest value are selected for next step. This cycle is repeated with 10 subsets until one of them gives the minimum [VV].

2.4. Case Study

In this study, a section of land located at Ondokuz Mayis University in Samsun province was chosen as the study area. A data set consisting of 411 points was used (Figure 2).

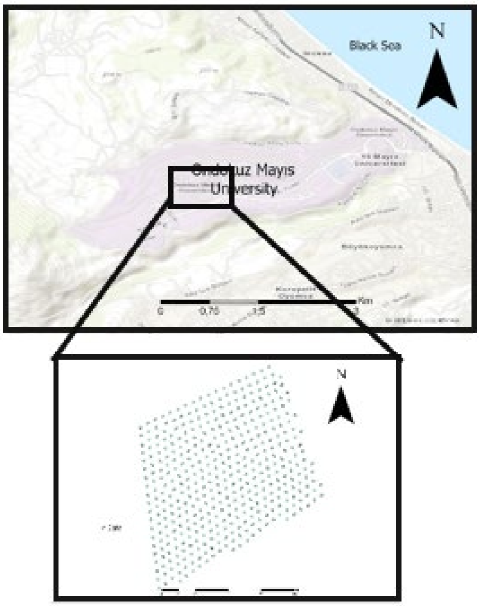

Figure 2. The study area and distrubition of 411 points

Points have X, Y and Z coordinates. By using these points, the polynomial surface is fitted with a 2nd degree polynomial. Then outlier measurement test was performed with three different methods.

3. Results

3.1. The Least Square Method

First, the LS method was tried. The method determined 14 out of 411 points as outlier measurement (Figure 3).

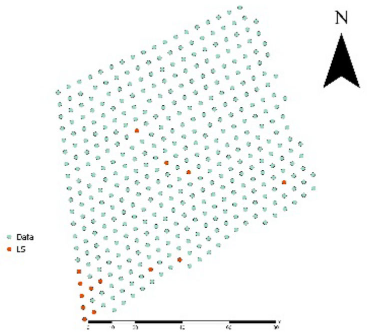

Figure 3. Outlier measurements of LS method

A compatible data group consisting of 397 points remained.

3.2. The Least Absolute Value Method

Then, LAV method was tried. The method determined 18 out of 411 points as outlier measurement (Figure 4).

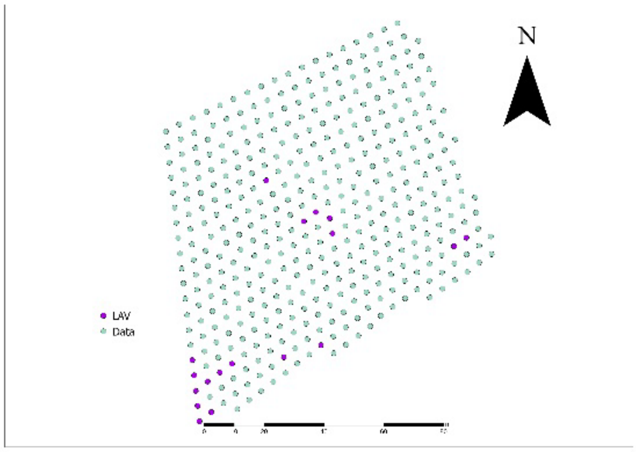

Figure 4. Outlier measurements of LAV method

A compatible data group consisting of 393 points remained.

3.3. The Least Trimmed Square Method

In LTS method, the trimming parameter h can be set between $\frac{n}{2} < h \leq n$. LS and LAV methods determined 14 and 18 points as outlier measurement. According to these numbers, h is set to 391. It means that 20 points could be outliers (Figure 5).

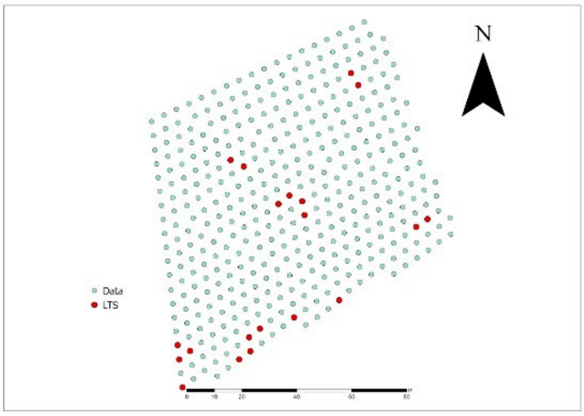

Figure 5. Outlier measurements of LTS method

4. Discussion and Conclusion

In this study, the outlier test performances of the LS, LAV and LTS methods are analyzed by fitting real data points to a surface with a 2nd degree polynomial. As a result, the LS and LAV methods found 14 and 18 outliers respectively. Considering outlier points of LS and LAV methods, 14 points of all 18 outliers found by LAD are common with those of LS (Figure 6).

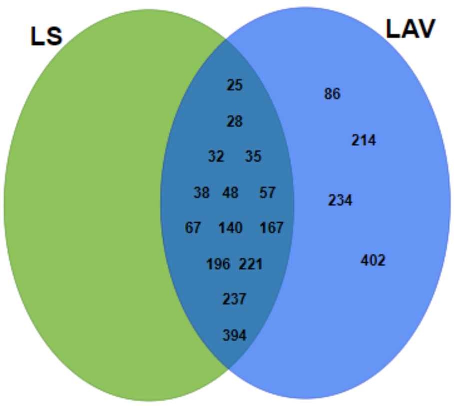

Figure 6. Venn diagram of LS and LAV methods.

LS and LTS have found 10 common outlier points which are less than LS and LAD (Figure 7).

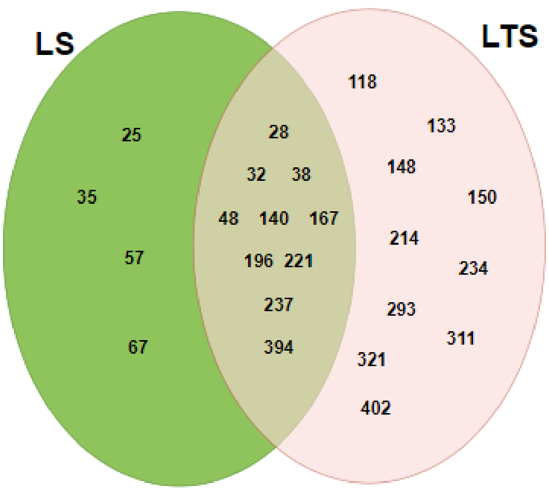

Figure 7. Venn diagram of LS and LTS methods.

Then venn diagram of all three methods are given in Figure 8.

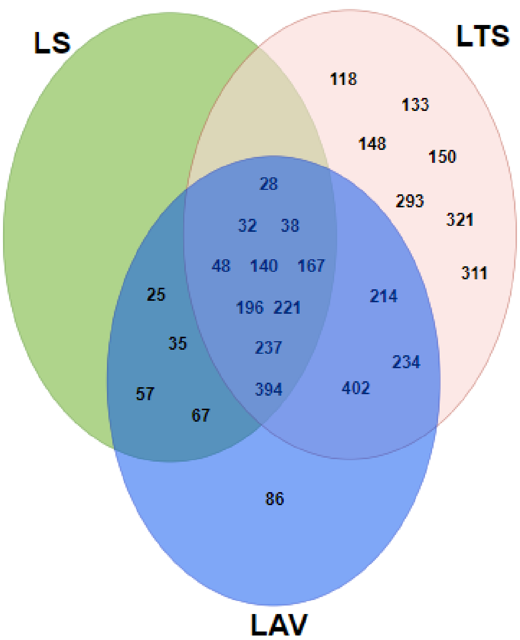

Figure 8. Venn diagram of LS, LAV and LTS methods.

The points determined as outlier measurements by the methods were removed from the cluster and RMSE was calculated (Table 1). Here, it is seen that the results are generally close to each other. But the smallest value was found in the LTS method. The largest value was found in the LS method.

Table 1. RMSE values of LS, LAV and LTS methods

Method	LS	LAD	LTS
RMSE (m)	0,1872	0,1829	0,1821

In this study, it is seen that the methods give close results compared to each other. Therefore, we can say that they can be used interchangeably. The study can be repeated by expanding the working area.

Author contributions

Ulku Kirici Yildirim and Hasan Dilmac worked together on methodology, writing, editing and review. **Yasemin Sisman** was the advisor of the study.

Conflicts of interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

Research and publication ethics were complied with in the study.

References

Atkinson, A. C., & Cheng, T. C. (1999). Computing least trimmed squares regression with the forward search. Statistics and Computing, 9(4), 251–263. https://doi.org/10.1023/A:1008942604045

Ayan T. (1992, Uyuşumsuz Ölçüler Testi, Harita ve Kadastro Mühendisliği Dergisi, 72, 1992,38-46

Baarda W (1968) A Testing Procedure for Use in Geodetic Networks. Publications on Geodesy, 2(5), 97.

Bektas S., Sisman Y., 2010, “The comparison of L1 and L2-norm minimization methods” International Journal of the Physical Sciences, 5(11), 1721 – 1727.

Berné Valero, J., & Baselga Moreno, S. (2005). Robust estimation in geodetic networks. Robust Estimation in Geodetic Networks, 17(17), 7–22. https://doi.org/10.5209/rev_FITE.2005.v17.12525

Dilmaç, H., & Şişman, Y. (2023). New approaches for outlier detection: The least trimmed squares adjustment. International Journal of Engineering and Geosciences, 8(1), 26-31

Erdogan, B. (2014): An Outlier Detection Method in Geodetic Networks Based on The Original Observations. Boletim de Ciencias Geodesicas, 20(3), 578–589. https://doi.org/10.1590/S1982-21702014000300033

Fang, X. (2015). Weighted total least-squares with constraints : a universal formula for geodetic symmetrical transformations. 459–469. https://doi.org/10.1007/s00190-015-0790-8

Ghilani, C. D. (2017). Adjustment computations: spatial data analysis. John Wiley & Sons.

Hampel, F., Ronchetti, E., Rousseeuw, P., and Stahel, W. 共 1986兲. “Robust statistics: The approach based on influence functions.” Wiley series in probability and statistics, Wiley, New York.

Hawkins, D. M. (1994). The feasible solution algorithm for least trimmed squares regression. Computational Statistics and Data Analysis, 17(2), 185–196. https://doi.org/10.1016/0167-9473(92)00070-8

Hekimoğlu, Ş. (2005). Do robust methods identify outliers more reliably than conventional tests for outliers? ZFV - Zeitschrift Fur Geodasie, Geoinformation Und Landmanagement, 130(3), 174–180.

Hekimoglu, S., & Erenoglu, R. C. (2009). New Method for Outlier Diagnostics in Linear Regression. Journal of Surveying Engineering, 135(3), 85–89. https://doi.org/10.1061/(asce)0733-9453(2009)135:3(85)

Huber, P. J. 共1981兲. “Robust statistics.” Wiley series in probability and statistics, Wiley, New York.

Kirici U., Sisman Y. (2015) The Determination Of The Best Fitting Geoid: A Case Study of Samsun, FIG Working Week, Bulgaria, Sofia.

Kirici U. (2016) Tusaga-Aktif (Cors-Tr) Noktaları İçin Farklı Amaç Fonksiyonlarına Göre Bölgesel Dönüşüm Katsayılarının Elde Edilmesi, Yüksek Lisans Tezi, Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü, Samsun.

Koch K R (1999) Parameter Estimation and Hypothesis Testing in Linear Models. 2nd Ed. Springer-Verlag, Berlin-Heidelberg, New York.

Koch, I., Veronez, M. R., da Silva, R. M., Klein, I., Matsuoka, M. T., Gonzaga, L., & Larocca, A. P. C. (2017). Least trimmed squares estimator with redundancy constraint for outlier detection in GNSS networks. Expert Systems with Applications, 88, 230–237. https://doi.org/10.1016/j.eswa.2017.07.009

Li, L. M. (2005). An algorithm for computing exact least-trimmed squares estimate of simple linear regression with constraints. Computational Statistics and Data Analysis, 48(4), 717–734. https://doi.org/10.1016/j.csda.2004.04.003

Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2014). On the least trimmed squares estimator. Algorithmica, 69(1), 148–183. https://doi.org/10.1007/s00453-012-9721-8

Pope A J (1976) The statistics of residuals and the detection of outliers. In NOAA Technical Report (p. 98).

Rousseeuw, P. J. (1984). Least median of squares regression. Journal of the American Statistical Association, 79(388), 871–880. https://doi.org/10.1080/01621459.1984.10477105

Rousseeuw P J, Leroy A M (1987) Robust Regression and Outlier Detection. John Wiley and Sons, New York.

Rousseeuw P J, Van Driessen K (2006) Computing LTS Regression for Large Data Sets. Data Min Knowl Disc 12, 29-45. https:/doi.org/10.1007/s10618-005-0024-4

Simkooei A.A., 2003, Formulation of L1 Norm Minimization in Gauss-Markov Models, J.Surv.Eng., 129(1):37-43.

Sisman Y.,2014, Coordinate transformation of cadastral maps using different adjustment methods, Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House,37-41 Mortimer Street, London W1T 3JH, UK., Çin.

Sisman, Y. (2010). Outlier measurement analysis with the robust estimation. Scientific Research and Essays, 5(7), 668–678.