"Environmental Engineering" 10th International Conference
Vilnius Gediminas Technical University
Lithuania, 27–28 April 2017
http://enviro.vgtu.lt

eISSN 2029-7092 / eISBN 978-609-476-044-0
Article ID: enviro.2017.250
DOI: https://doi.org/10.3846/enviro.2017.250

Subjective Breakdown Points of R-estimators Applied in Deformation Analysis

Patrycja Wyszkowska[1], Robert Duchnowski[2]

Institute of Geodesy, University of Warmia and Mazury, Olsztyn, Poland
E-mails: [1]pwyszkowska@wp.pl (corresponding author); [2]robert.duchnowski@uwm.edu.pl

Abstract. This paper presents practical aspect of the breakdown point theory in deformation analysis by applying R-estimators. The main aim of the paper is to determine impact of the probability of positive (or negative) gross errors and the number of such errors on the value of breakdown point of the estimates applied. Authors consider two types of networks: a levelling network and a horizontal one. Calculations are made for two cases, namely when observations are affected by gross errors in both measurement epochs or only in the second epoch. The main results are based on the Monte Carlo method, which is a very useful tool to solve such a geodetic problem. The simulations show that the breakdown point depends on the probability of positive gross errors but also on the number of epochs in which the gross errors occur. This is especially vivid in the case of levelling networks. Another interesting finding is that even if the number of gross errors exceeds the breakdown point, we can still get reasonable results; however, not always. Thus, the paper shows the probabilities that the method breaks down for several different cases. The paper includes some numerical tests, which provided practical information about the subjective breakdown points and their importance for R-estimates applied in deformation analysis.

Keywords: breakdown point, R-estimation, point displacement, Monte Carlo simulations.

Conference topic: Technologies of geodesy and cadastre.

Introduction

The theory of breakdown points is well known in robust statistics since the paper of Hampel (1971). On one hand, the breakdown point is an informative measure of robustness of a specific method; on the other, it has also some practical applications (see, e. g., Xu 2005). In general, the breakdown point gives information how many outliers must make a method to break down (or how many outliers a method can withstand before it breaks down). Thus, if we consider values of the breakdown points, we can compare two (or more) robust methods with each other and chose the most appropriate to the current problem or task. We can also expect when the method provides acceptable results, and when it must break down. However, there are some special cases when the method can withstand more outliers that would result from the value of the breakdown point. Note that in such cases, we cannot be sure that results of estimation are "good" but we can compute the probability of such acceptable results. In some cases such probability can be surprisingly high (see, Xu 2005). Such computations must be based on additional knowledge about distribution of outliers, and/or the experience of the analyst. For that reason, Xu (2005) called them probabilities of subjective breakdown points. In such a context, the subjective breakdown point may achieve values bigger than 0.5 (or 50%), which is the biggest value for traditional breakdown points. The subjective breakdown points and their corresponding probabilities describe how the method responds for bigger number of outliers, which is interesting from the theoretical point of view. On the other hand, such knowledge might also have practical importance, for example, when one can predict the locations of outliers. There are several robust methods which are applied in geodesy of surveying, like for example, robust M-estimation (with several different methods), M_{split} estimation or R-estimation (for example, Xu 1989, 2005; Yang 1994; Kargoll 2005; Duchnowski 2010; Wiśniewski 2009, 2014; Nowel, Kamiński 2014). In the present paper, we will focus on R-estimates which can be applied, for example, in deformation analysis (Duchnowski 2009, 2010, 2013; Kargoll 2005). The chosen R-estimates, namely Hodges-Lehmann estimates (HLE), were the basis for the method of deformation analysis which was proposed by Duchnowski (2010). Note that, such method was especially recommended for testing stability of potential reference points (PRPs). The robustness of such an approach was discussed by Duchnowski (2011). It was shown that there might be two sources of outliers when HLEs are applied in deformation analysis. The first one is obvious, namely gross errors. The second type of outliers results from instability of other points in network (or instability of other possible reference points). If outliers of such two types occur in one network, the percentage of such observations might be bigger that the breakdown point, hence the method must break down. However, some simulations show that even then the final results may be acceptable in many practical cases (for example, Duchnowski, Wiśniewski 2014). For such reason, it would be interesting to analyse the subjective breakdown points of HLEs. Thus, the paper will investigate such breakdown point on the example of levelling or horizontal networks. It will also present computation of probabilities of subjective breakdown points for different values of assumed probability of positive gross errors for several variants. All the results will be based on Monte Carlo simulations.

R-estimates in deformation analyses

The first rank-based estimates (R-estimates) were proposed by Hodges and Lehmann (1963). The authors considered two general cases: one-sample problem, namely estimation of the expected value and two-sample problem, estimation of the shift between two samples. Both such variants are known as Hodges-Lehmann's estimates (HLEs). From the practical point of view, the estimate of the shift is more useful in the case of geodetic applications, especially in deformation analyses (see, for example, Duchnowski 2009). Thus, let us consider two-sample problem. First, let us assume that random variables X_j $(1 \leq j \leq m)$ and also Y_i, $(1 \leq i \leq n)$ are independently and identically distributed with the respective continuous distributions F(x) or G(y) = F(x−Δ). Now, we can consider two independent samples $x_1, x_2, \ldots, x_m$ and $y_1, y_2, \ldots, y_n$ which are sets of realizations of the mentioned variables. Hodges and Lehmann (1963) applied the Wilcoxon test statistic and they proposed well-known form of HLE of the shift Δ

$$\hat{\Delta}^{\mathrm{HL}} = \mathrm{med}(y_{\mathrm{i}} - x_{\mathrm{j}}), \quad (1)$$

where: $\hat{\Delta}^{HL}$ – HLE of the shift; med – median operator.

Considering application in deformation analyses, that estimate is a natural estimate of a vertical (or horizontal) displacement of a network point if only the samples in question are sets of point coordinates (*X, Y, Z* or *H* respectively) at two different measurement epochs. The estimate of Eq. (1) is a good alternative for more conventional types of estimation, like for example the least squares estimates (LSE) or M-estimates (see, Duchnowski, Wiśniewski 2014, 2016). In some cases, one should assume the initial values of the point coordinates or apply the initial values of residuals instead of point coordinates (Duchnowski 2010; Cymerman *et al.* 2016); however such approaches are not used in the present paper.

In the case of levelling or horizontal networks, the coordinates which are elements of the samples $x_1, x_2, \ldots, x_m$ (or $y_1, y_2, \ldots, y_n$) are computed separately for each measurement epoch. The computations are based on the coordinates of the reference points and the measurements from the respective epoch. The coordinates should be computed in all possible independent ways (see, Duchnowski 2013). Thus, when one computes all the coordinates for two epochs then the shifts of the point coordinates can be estimated by applying Eq. (1) and the vertical (or horizontal) point displacement can be obtained.

Subjective breakdown points of HLE

Consider the following types of networks: two levelling networks and one horizontal network. Calculations are made for two cases, when outliers occur only in the second measurement epoch or in both measurement epochs. The absolute values of gross errors vary from 0.015 m to 0.050 m for height differences and for distances, and from 0.015^g to 0.050^g as for horizontal angles (the absolute values of such errors are simulated assuming that they are uniformly distributed within such intervals). The signs of the gross errors are randomly chosen, and the probability of positive (or negative) gross errors varies from 0.5 to 1.0 (simulating the prior information about the properties of such errors). All results presented in this section are based on Monte Carlo simulations. Without loss of generality we assumed that all network points are stable, hence from the theoretical point of view all estimated displacements are equal to zero.

Levelling networks

The first simulated levelling network consists of three reference points and one object point. There are 3 observations (height differences) between each reference point and the object point. The observations are simulated for two measurement epochs. Considering the properties of R-estimates (Duchnowski 2011), the method cannot be disturbed by a single outlier. Now let us consider the case when there are two outlying observations. Figure 1 presents the probability of the breakdown of R-estimate of point displacement when there are 2 outliers among 3 observations and for several probabilities of positive outliers. When 2 outliers occur only in second measurement epoch, then the probability of the breakdown of R-estimate increases with the increase of the probability of positive outliers. When the probability of positive outliers equals 1, the tested R-estimate certainly breaks down. On the other hand, the probability of the breakdown decreases with the increase of the probability of positive outliers in the case of 2 outliers in both measurement epochs. Thus, when probability of positive outliers is high, one should expect smaller probability of the breakdown of R-estimate when 2 outliers occur in both measurement epochs.

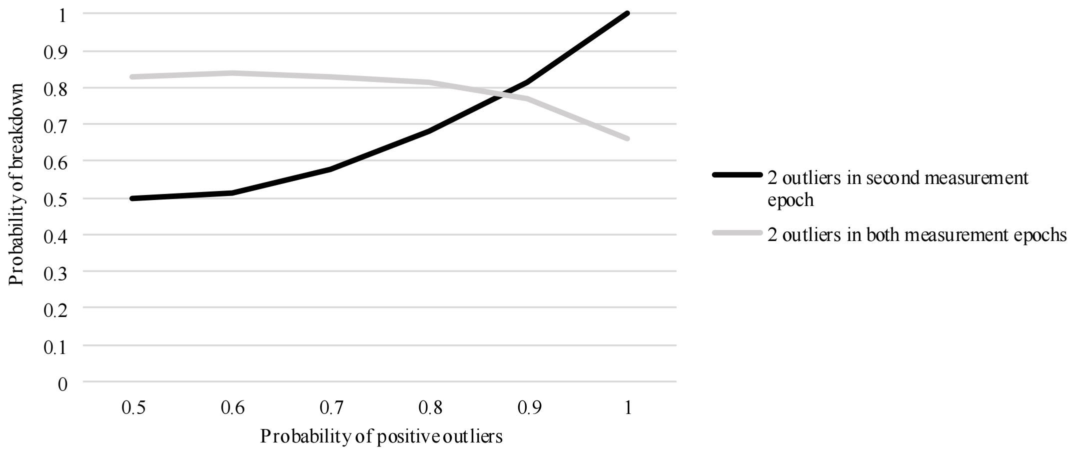


Fig. 1. Probability of the breakdown of R-estimate of point displacement in levelling network with three reference points and one object point

The second simulated levelling network contains five reference points and one object point. There are 5 observations (height differences) between each reference point and the object point in this network. Figure 2 shows the probability of the breakdown of R-estimate of point displacement in that levelling network for several different probabilities of positive outliers and when the outliers occur only in the second measurement epoch. In case of 1 or 2 outliers of 5 observations probability of the breakdown of R-estimate of point displacement equals 0 for every probability of positive outliers. The method can deal with such small number of outliers, which in fact is smaller than the classical method's breakdown point (see, Duchnowski 2011). When there are 3 outliers, R-estimate breaks down but with different probability for different probabilities of positive outlier. When the probability of positive outliers equals 1, the probability of breakdown of R-estimate also equals 1. But if the probability of positive gross errors is closer to 0.5, then the estimation can withstand 3 outliers with rather big probability. Generally speaking, the bigger probability of positive outliers, the larger probability that R-estimate breaks down. Besides, the probability of breakdown of R-estimate increases with bigger number of outliers.

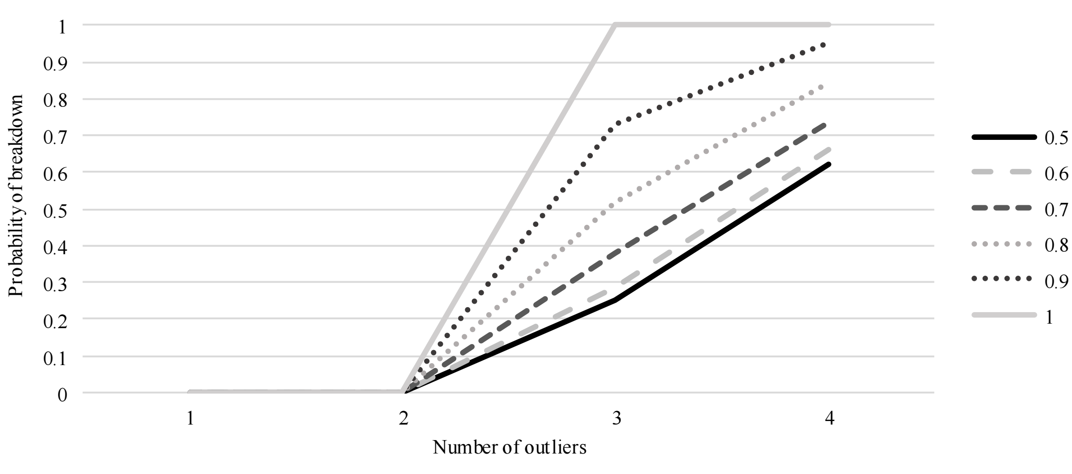


Fig. 2. Probability of the breakdown of R-estimate of point displacement in levelling network with five reference points and one object point for several probabilities of positive outliers and when the outliers occur only in the second measurement epoch

Figure 3 presents the second possible case, namely when the outliers occur in both measurement epochs. The probability of the breakdown of R-estimate of point displacement equals 0 for every probability of positive outliers, when there is only 1 outlier (among 5 observations) in each of the measurement epochs. In case of 2 outliers, the probability of the breakdown of R-estimate of point displacement equals 0 only when probability of positive outliers equals 1. Thus, in such case the method can withstand the outliers. Overall, the probability that R-estimate breaks down increases with the increase of the number of outliers. What is more, the bigger probabilities of breakdown are obtained for the smaller probabilities of positive gross errors. Thus, such an effect is opposite to that presented in Figure 2, when outliers occur only in the second epoch.

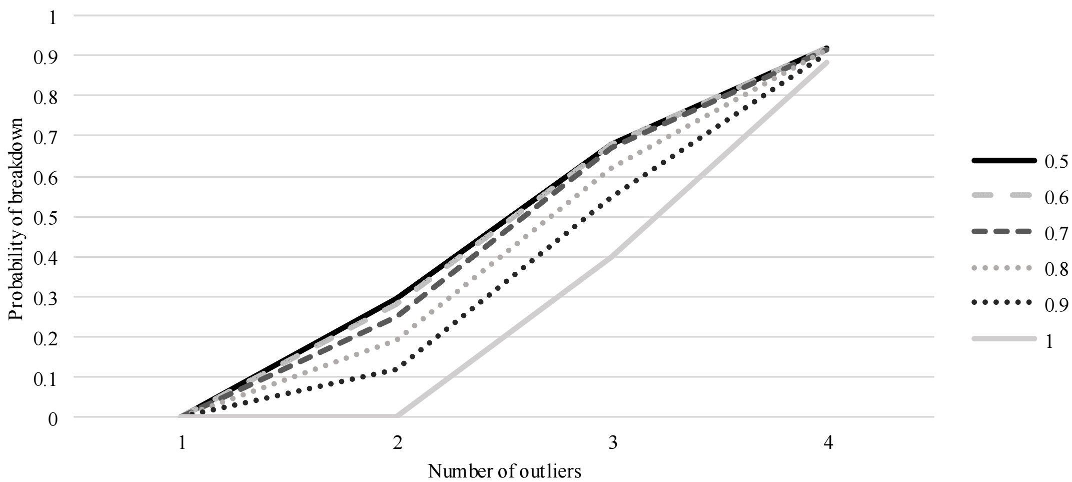
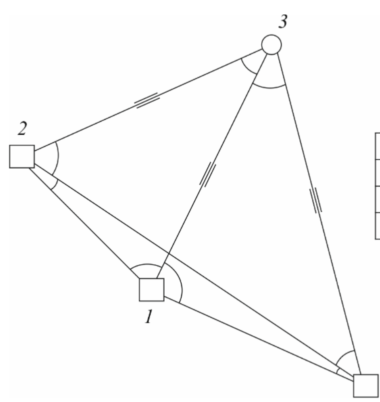


Fig. 3. Probability of the breakdown of R-estimate of point displacement in levelling network with five reference points and one object point for several probabilities of positive outliers and when the outliers occur in both measurement epochs

Horizontal network

Let us consider a horizontal network with three reference points and one object point. We assume two measurement epochs and the following independent observations: 6 distances and 8 horizontal angles (Fig. 4). The issue is to estimate horizontal displacement of the point 3, thus to apply R-estimate, one should compute the horizontal coordinate of that point in all possible independent ways. Here, one can apply: six combined sections (based on the reference points *2 - 4*, *2 - 1*, *4 - 2*, *4 - 1*, *1 - 4*, *1 - 2*) and one angular resection (*2 - 1 - 4*). One can use such sets of computed coordinates to calculate R-estimate of Eq. (1) and obtain the shift of the point according to the axes *X* or *Y*, respectively. Here, we will present only results obtained for the coordinate *X* (the results for the other coordinate are very similar).



	X [m]	*Y* [m]
1	-100.00	0.00
2	0.00	-100.00
3	100.00	100.00
4	-200.00	200.00

Fig. 4. Simulated horizontal network

Figure 5 presents the probability of the breakdown of R-estimate in simulated horizontal for several values of probability of positive outliers and when the outliers occur only in the second measurement epoch. The first conclusions are rather obvious: the probability of the breakdown of HLE equals 0 if there are not more than 3 outliers of all 14 observations, and in the case of 13 (or 14) outliers the probability of the breakdown of R-estimate equals 1. They both follow from the general properties of HLE. Note that there are 14 measurements at each of the epochs but we compute "only" 7 different values of coordinates of the point 3. Thus, if there are 4 or more outliers among the measurements then probably more than half values of the computed coordinates in the second epoch are outlying. The method cannot withstand 4 (or more) outliers; however probability of breaking down of HLE is rather small and smaller than 0.2 (for all considered probabilities of positive outliers). One can also see smaller differences among the results obtained for different values of probability of positive outliers, which differs from the findings concerning the levelling network.

The maximum differences are about 0.3 and concern the variants with 6, 7 or 8 outliers. Note that such differences are even smaller when the outliers occur at both measurement epochs (see, Fig. 6). Thus, we can conclude that in the case of horizontal networks, the knowledge of the particular probability of a positive outlier is not so important for the final results of the estimation.

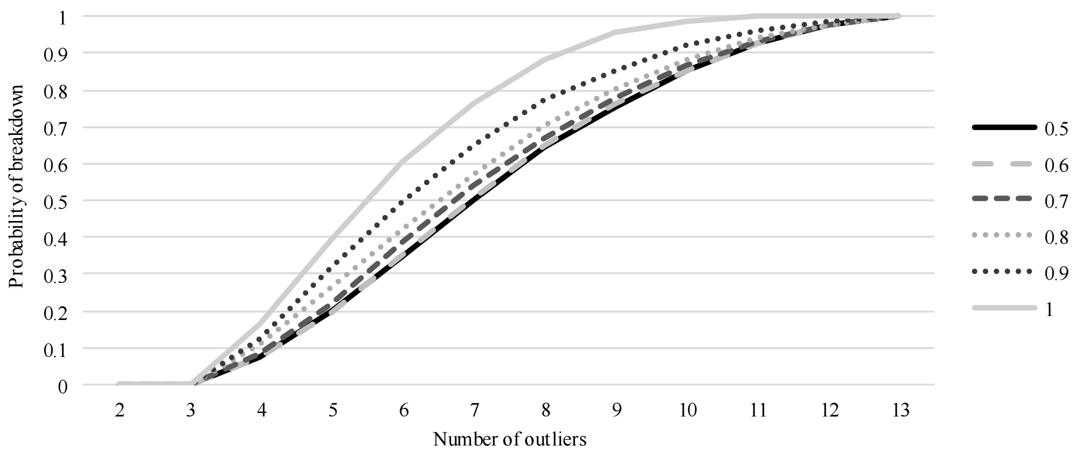


Fig. 5. Probability of the breakdown of R-estimate of point displacement in horizontal network (change of the coordinate *X*) for several probabilities of positive outliers and when the outliers occur only in the second measurement epoch

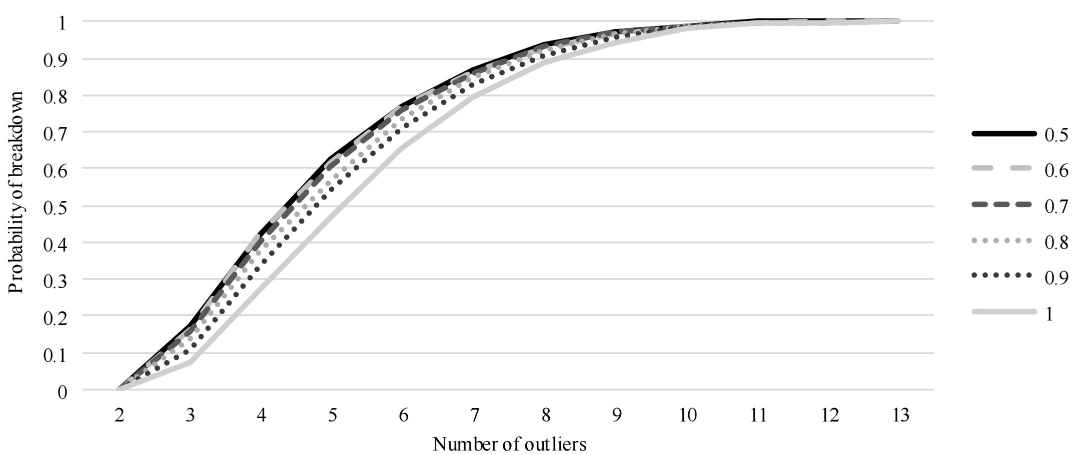


Fig. 6. Probability of the breakdown of R-estimate of point displacement in horizontal network (change of the coordinate X) for several probabilities of positive outliers and when the outliers occur in both measurement epochs

Hence, the subjective breakdown points have less importance in such cases. To explain such a finding, one should realize that the positive gross error of a particular measurement, namely distance or angle, does not mean that the coordinate computed on the base of such outlying observation will be considered as a positive outlier (an observation much bigger than "good" observations) among other computed values of coordinates. Generally speaking, it is hard to predict location of outliers among the computed coordinates if we know only the sign of the gross errors of measurements (note that in the case of levelling networks the situation is much simpler and just obvious). It is also worth noting that in both types of geodetic networks considered here one can notice the same influence of occurrence of outliers at one or both measurement epochs. In the first case, the probability of breaking down is smaller when the probability of a positive outlier is closer to 0.5. In the second one, the situation is exactly the opposite.

Conclusions

The paper presents some considerations concerning the breakdown points of HLE of the shift, which can be applied in deformation analyses. The main objective was to examine the subjective breakdown points, namely the situations when the number of outliers exceeds the traditional breakdown point. The obtained results show that if the number of outliers is only a little bigger that the breakdown point then the probability of subjective breakdown points is relatively small (< 0.3). Such probability depends on the number of outliers but also on the occurrence of outlying observations (whether they occur at one or both measurement epochs). The prior information about the sign of the gross errors, namely the probability of positive or negative gross errors, is especially important for levelling networks. In such a case, the probability of the subjective breakdown points depends largely on such a probability. In the case of horizontal networks such an influence is much less significant, which results from the relation between a particular gross error of measurement and the value of the coordinate computed by applying such an outlying measurement.

The results presented in the paper show that HLE can give "good" results even if the percentage of outliers exceeds the breakdown points. However, it is not the only practical conclusion. One should note that we assume that the sign of the gross error is chosen randomly (but with assumed probability). However, in deformation analyses we sometimes have prior information about outliers, for example, when we test stability of PRPs. In such a case, some outliers might result from instability of other PRPs (not from gross errors), thus their signs are conditionally predictable (Duchnowski 2010, 2011). In such a context, the knowledge of the subjective breakdown points seems especially important because it give us information when the method breaks down (or in other words, how many unstable PRPs the method can withstand). The paper shows that such a conclusion concerns especially levelling networks. Another important problem arises when two types of outliers in question occur together. Duchnowski (2011) partly addressed such a problem; however, this surely requires further research also in the context of subjective breakdown points.

Funding

This work was supported by Institute of Geodesy, University of Warmia and Mazury in Olsztyn [the statutory research no. 28.610.002-300].

References

Cymerman, M.; Duchnowski, R.; Kopiejczyk, A. 2016. Selection of initial parameters in R-estimated applied to deformation analysis in leveling networks, *Journal of Surveying Engineering* 142(2). https://doi.org/10.1061/(ASCE)SU.1943-5428.0000151

Duchnowski, R. 2009. Geodetic application of R-estimation – levelling network examples, *Technical Sciences* 12: 135–144.

Duchnowski, R. 2010. Median-based estimates and their application in controlling reference mark stability, *Journal of Surveying Engineering* 136(2): 47–52. https://doi.org/10.1061/(ASCE)SU.1943-5428.0000014

Duchnowski, R. 2011. Robustness of strategy for testing levelling mark stability based on rank tests, *Survey Review* 43(323): 687–699. https://doi.org/10.1179/003962611X13117748892551

Duchnowski, R. 2013. Hodges-Lehmann estimates in deformation analyses, *Journal of Geodesy* 87(10): 873–884. https://doi.org/10.1007/s00190-013-0651-2

Duchnowski, R.; Wiśniewski Z. 2016. Accuracy of the Hodges-Lehmann estimates computed by applying Monte Carlo simulations, *Acta Geodaetica et Geophysica.* https://doi.org/10.1007/s40328-016-0186-0

Duchnowski, R.; Wiśniewski, Z. 2014. Comparison of two unconventional methods of estimation applied to determine network point displacement, *Survey Review* 46(339): 401–405. https://doi.org/10.1179/1752270614Y.0000000127

Hampel, F. R. 1971. A general qualitative definition of robustness, *The Annals of Mathematical Statistics* 42(6): 1887–1896. https://doi.org/10.1214/aoms/1177693054

Hodges, J. L. Jr.; Lehmann, E. L. 1963. Estimates of location based on rank tests, *The Annals of Mathematical Statistics* 34(2): 598–611. https://doi.org/10.1214/aoms/1177704172

Kargoll, B. 2005. Comparison of some robust parameter estimation techniques for outlier analysis applied to simulated GOCE mission data, in *GGSM 2004 IAG International Symposium*, 30 August–3 September 2004, Porto, Portugal: 77–82. https://doi.org/10.1007/3-540-26932-0_14

Nowel, K.; Kamiński, W. 2014. Robust estimation of deformation from observation differences for free control networks, *Journal of Geodesy* 88(8): 749–764. https://doi.org/10.1007/s00190-014-0719-7

Wiśniewski, Z. 2009. Estimation of parameters in a split functional model of geodetic observations (M_{split} estimation), *Journal of Geodesy* 83(2): 105–120. https://doi.org/10.1007/s00190-008-0241-x

Wiśniewski, Z. 2014. M-estimation with probabilistic models of geodetic observations, *Journal of Geodesy* 88(10): 941–957. https://doi.org/10.1007/s00190-014-0735-7

Xu, P. 1989. On robust estimation with correlated observations, *Bulletin Géodésique* 63(3): 237–252. https://doi.org/10.1007/BF02520474

Xu, P. 2005. Sign-constrained robust least squares, subjective breakdown point and the effect of weights of observations on robustness, *Journal of Geodesy* 79(1): 146–159. https://doi.org/10.1007/s00190-005-0454-1

Yang, Y. 1994. Robust estimation for dependent observation, *Manuscripta Geodaetica* 19: 10–17.