A Case Study for Outlier Detection Efficiency Based on M estimations of Different Weight Functions and Models

TU Xian-qin[1,2,a], YI Dong-yun[1] and ZHOU Hai-yin[1]

[1]College of Science, National University of Defense Technology, Changsha, China

[2]PLA 91550 Unit, Liaoning Dalian, China

[a]tuxq79@gmail.com

Keywords: Outlier detection, Robust M-estimation, Tracking Data modeling

Abstract. The effects of the model and weight function on outlier detection are evaluated by the simulated optical and radar observations. The iterative reweighted M-estimation based on different iterative reweighted functions is used for the outlier detection test. Three typical models of the optical and radar tracking data are compared for their effect on the outlier test. The simulated results show that different weight functions have small difference on the outlier detection efficiency and a good modeling selection for the same dataset is an key factor for a best outlier detection procedure.

Introduction

To date many approaches have been deeply developed to identify the outliers more accurately. There are two different strategies to mitigate the presence of outliers[1][2]. The first is to identify outliers using outlier tests, and then reject the observation exceeding the critical value for the desired significance level based on the statistic test. If multiple outliers exist then the single outlier test is applied iteratively with the strategy of removing the largest observation first until all the outliers have been removed. The second method is to use robust methods that without removing any observations but down weight suspect observations. When multiple outliers exist, the first method often failed because of the separation. Separability refers to the ability to distinguish or separate outliers from the other normal observations. The poorly separated observations adversely affect the solution of the system by manifesting a high risk of incorrectly flagging a 'good' observation as an outlier or vice versa.

In the methods of the second kind, the M-estimation, which is implemented by the iterative reweighted least square, is popular for its simplicity. Different weight functions have been defined for the M-estimation method; if this is the case then which proper weight function should be used for the outlier detection of the given problem? Ideally, the method chosen should be capable of handling multiple outliers. The method should also be resilient to the effects of incorrect exclusion where only some of the outliers are identified and wrong exclusion where a correct observation is identified. If neither incorrect exclusion nor wrong exclusion occurs then it is a correct exclusion as all of the outliers and only the outliers have been excluded.

As a result of this, it is the intention of the study to compare the abilities of the outlier test to correctly exclude outliers in three typical models of the optical and radar measurements. The comparison is based on miss rate and false alarm rate. The model effect to outlier detection is also considered in the comparison.

Methodology

Three typical models are considered in the research. All of the models use the same set or subset of the dataset. Different models are used to evaluate for the model effect in the outlier detection.

Single Epoch Model. The first model is called single epoch model which uses multiple synchronized observations from different measure devices. The system measurement equation is

made by the synchronized measurements of different optical or radar measure devices from the same epoch time.

The measurement equations of the optical or radar measure devices can be written by a general non-linear vector function as:

$$\boldsymbol{R} = \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{h}(\boldsymbol{x})$ is the non-linear measurement equations; $\boldsymbol{\varepsilon}$ is the observation errors vector.

The linearized measurement equations at the preliminary position $\boldsymbol{x}_0$ of the tracking target is

$$\Delta\boldsymbol{R} = \boldsymbol{H} \cdot \Delta\boldsymbol{x} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\Delta\boldsymbol{R}$ is the OC (observation minus the computed) residuals vector; $\boldsymbol{H}$ is the linearize partial derivatives matrix of $\boldsymbol{h}(\boldsymbol{x})$ with respect to $\boldsymbol{x}$ at $\boldsymbol{x} = \boldsymbol{x}_0$, $\boldsymbol{H} = \left.\dfrac{\partial \boldsymbol{h}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}_0}$; $\Delta\boldsymbol{x}$ is the estimated parameters vector of the equation.

Time Series Model. In the second model, which is called the time series model in the paper, is constructed by time series data of a selected measure element type from one measure device. The time series data $R(t)$ is modeled by a polynomial:

$$R(t) = \sum_{j=0}^{q} a_j t^j + \varepsilon(t), \tag{3}$$

where a_j is the coefficients of the polynomial; q is the order of the polynomial; $\varepsilon(t)$ is the observation error at time t. There is numerical stability problem for the applications using the Eq. 1. So the orthogonal polynomials replace the simple one in the Eq. 1:

$$p_j(t) = t^j + \sum_{k=0}^{j-1} a_{k,j} t^k, \ j = 0,1,\cdots,q, \tag{4}$$

where $\left\{ p_j(t) \right\}$ satisfy the following equations:

$$\sum_{i=1}^{m} p_l(t_i)\, p_k(t_i) = \begin{cases} 0, & k \neq l, \\ \sum_{i=1}^{m} p_l^2(t_i), & k = l. \end{cases} \tag{5}$$

Then get the discrete version of the Eq. 3 at time t_i:

$$R(t_i) = \sum_{j=0}^{q} \beta_j p_j(t_i) + \varepsilon(t_i), \tag{6}$$

where β_j is the coefficients of the orthogonal polynomials, $\varepsilon(t_i)$ is the observation error at time t_i. The vectorized version of the Eq. 6 is :

$$\boldsymbol{R} = \boldsymbol{C} \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{7}$$

where $\boldsymbol{R} = \begin{pmatrix} R(t_1) & R(t_2) & \cdots & R(t_m) \end{pmatrix}^T$, $\boldsymbol{C}$ is the known matrix, $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 & \beta_1 & \cdots & \beta_q \end{pmatrix}^T$ is the estimated polynomial coefficients vector, $\boldsymbol{\varepsilon}$ is the errors vector, m is the total sampling count.

Fusion Model. The third model called the fusion model. The position of the tracking target is defined as:

$$\boldsymbol{x} = \begin{pmatrix} x & y & z \end{pmatrix}^T, \tag{8}$$

and $x(t_i) = \sum_{j=0}^{q} \beta_{x,j} p_j(t_i)$, $y(t_i) = \sum_{j=0}^{q} \beta_{y,j} p_j(t_i)$, $z(t_i) = \sum_{j=0}^{q} \beta_{z,j} p_j(t_i)$, where $\beta_{.,j}$ is the coefficients of the orthogonal polynomials.

From Eq.1 and Eq. 8, we can get the following equations, including a linearized version:

$$\boldsymbol{R} = \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{\varepsilon} = \boldsymbol{g}(\boldsymbol{\beta}) + \boldsymbol{\varepsilon}, \tag{9}$$

$$\Delta\boldsymbol{R} = \boldsymbol{G} \cdot \Delta\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{10}$$

where $\Delta \boldsymbol{R}$ is the OC residuals vector of multiple measure devices and multiple observation epochs; $\boldsymbol{G}$ is the linearized partial derivatives matrix; $\Delta \boldsymbol{\beta}$ is the estimated vector of the equation.

The above equations of the three types can be solved by the robust M-estimation introduced by Huber[1]. M-estimation is a generalized form of maximum likelihood estimation (MLE) and an iteratively reweighted LS estimation is used to solve it. Several weight functions were introduced in the past few years[2-6]. The weight functions considered in the paper are listed in the following Table 1.

Table 1. The weight functions used in M-estimation

Method Name	Weight function	Condition	Critical Value
Huber	1	$\lvert v\rvert \le c$	$c = 1.5 \sim 2.0$
	$c/\lvert v\rvert$	$\lvert v\rvert > c$	
Hampel	1	$\lvert v\rvert < a$	$a = 0.4 \sim 1.5$
	$a/\lvert v\rvert$	$a \le \lvert v\rvert < b$	$b = 0.8 \sim 2.5$
	$a(c-\lvert v\rvert)/((c-b)\lvert v\rvert)$	$b \le \lvert v\rvert < c$	$c = 2.0 \sim 5.0$
	0	$\lvert v\rvert \ge c$	
Andrews	$(\lvert v\rvert/c)^{-1}\sin(\lvert v\rvert/c)$	$\lvert v\rvert \le c\pi$	$c = 1.5 \sim 2.0$
	0	$\lvert v\rvert > c\pi$	
Tukey	$\left(1-(\lvert v\rvert/c)^2\right)^2$	$\lvert v\rvert \le c$	$c = 4.0 \sim 6.0$
	0	$\lvert v\rvert > c$	
IGGIII	1	$\lvert v\rvert \le c_0$	$c_0 = 2.0 \sim 3.0$
	$c_0/\lvert v\rvert$	$c_0 < \lvert v\rvert \le c_1$	$c_1 = 4.5 \sim 8.5$
	0	$\lvert v\rvert > c_1$	
Danish	1	$\lvert v\rvert \le c$	$c = 1.5 \sim 2.0$
	$\exp\left(-(\lvert v\rvert/c)^2\right)$	$\lvert v\rvert > c$	
L1-norm	$1/\lvert v\rvert$	no	no

Results and Analysis

Outliers of varying size and number were added into the simulated observations according to an predefined probability. The three standard deviation threshold is chosen in the outlier detection test. It is most likely that robust methods would have significantly reduced the influence of residuals of this magnitude. Once a residual greater than three standard deviations, it is identified as an outlier.

The same simulated dataset was used for the three typical model. The comparisons of different models for different weight functions are displayed in the Table 2, 3 and 4 respectively.

Table 2. The outlier detection efficiency of the single epoch synchronized model

Weight function	Miss rate	False alarm rate
Huber	0.0972	0.0016
Hampel	0.0555	0.0118
Andrews	0.1111	0.0032
Tukey	0.1111	0.0049
IGGIII	0.1250	0.0021
Danish	0.0972	0.0048
L1-norm	0.0555	0.0061

Table 3. The outlier detection efficiency of the time series model

Weight function	Miss rate	False alarm rate
Huber	0.0000	0.0010
Hampel	0.0000	0.0030
Andrews	0.0000	0.0018
Tukey	0.0000	0.0021
IGGIII	0.0000	0.0008
Danish	0.0000	0.0029
L1-norm	0.0000	0.0013

Table 4. The outlier detection efficiency of the fusion model

Weight function	Miss rate	False alarm rate
Huber	0.0000	0.0366
Hampel	0.0000	0.0423
Andrews	0.0000	0.0353
Tukey	0.0000	0.0379
IGGIII	0.0000	0.0344
Danish	0.0000	0.0423
L1-norm	0.0000	0.0381

From the above tables, it can be seen that the single epoch synchronized model has the largest miss rate and the fusion model has the largest false alarm rate. The time series model has the best successful rate among all the models. The results showed small difference on different weight functions, which can be neglectable.

References

[1] P. J. Huber, E. M. Ronchetti. Robust Statistics, Second Edition [M]. Hoboken, New Jersey: John Wiley & Sons, Inc., (2009).

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

[3] E. Gokalp, O. Gungor, Y. Boz. Evaluation of Different Outlier Detection Method for GPS Networks [J]. Sensors. (2008), 8(11): 7344–7358.

[4] Y. Boz, E. Gokalp. Robust Estimation of the Outliers in GPS Baseline Components [C]. In Shaping the Change XXIII FIG Congress. Munich, Germany, October 8-13 (2006).

[5] S. Hekimoglu. R. C. Erenoglu. Effect of heteroscedasticity and heterogeneousness on outlier detection for geodetic networks [J]. Journal of Geodesy. (2007), 81:137-148.

[6] Y. Yang. Robust estimation of geodetic datum transformation [J]. Journal of Geodesy. (1999), 73(5): 268–274.