

## OUTLIER DETECTION IN PARTIAL ERRORS-IN-VARIABLES MODEL

### *Detecção de outliers usando um modelo de observações de erro*

Jun Zhao<sup>1,2,3</sup>  
Qingming Gui<sup>4</sup>

<sup>1</sup>Xi'an Technical Division of Surveying and Mapping, 710054, Xi'an, China

<sup>2</sup>State Key Laboratory of Geodesy and Earth's Dynamics, 430077, Wuhan, China

<sup>3</sup>School of Surveying mapping, Information Engineering University, 450001, Zhengzhou, China. Email: zhaojun4368@163.com

<sup>4</sup>Institute of Science, Information Engineering University, 450001, Zhengzhou, China. Email: [guiqingming@126.com](mailto:guiqingming@126.com)

#### **Abstract:**

The weighed total least square (WTLS) estimate is very sensitive to the outliers in the partial EIV model. A new procedure for detecting outliers based on the data-snooping is presented in this paper. Firstly, a two-step iterated method of computing the WTLS estimates for the partial EIV model based on the standard LS theory is proposed. Secondly, the corresponding  $w$ -test statistics are constructed to detect outliers while the observations and coefficient matrix are contaminated with outliers, and a specific algorithm for detecting outliers is suggested. When the variance factor is unknown, it may be estimated by the least median squares (LMS) method. At last, the simulated data and real data about two-dimensional affine transformation are analyzed. The numerical results show that the new test procedure is able to judge that the outliers locate in  $x$  component,  $y$  component or both components in coordinates while the observations and coefficient matrix are contaminated with outliers.

**Keywords:** Partial EIV model; Two-step iterated method; Weighted total least-squares; Outlier detection; Data-snooping; Two-dimensional affine transformation

#### **Resumo:**

O estimador dos Mínimos Quadrados Total é muito sensível à presença de outliers no modelo de observações de erro. Neste trabalho apresenta-se um novo modelo para detecção de outliers baseado na técnica *data-snooping*. Primeiro, é proposto um método iterativo para determinar o estimador dos Mínimos Quadrados Total na teoria dos Mínimos Quadrados. Em seguida, o teste estatístico  $w$  é construído para detectar outliers enquanto as observações e a matriz de coeficientes são contaminadas com a presença de outliers, sendo sugerido um algoritmo específico para detecção de outliers. Quando o fator de variância é desconhecido, ele deve ser estimado pelo método dos Mínimos Quadrados Medianos. Foram analisados dados simulados e reais. Os resultados numéricos mostraram que o método proposto é capaz de identificar se os

outliers se encontram nas componentes em  $x$  ou em  $y$ , enquanto as observações e a matriz de coeficientes são contaminados com outliers.

**Palavras-chave:** Modelo EIV Parcial; Método Iterativo Two-step; Estimador dos Mínimos quadrados Total; Detecção de Outlier; Data-snooping; Transformação Afim bidimensional.

## 1. INTRODUCTION

Gauss-Markov (G-M) model and least-squares (LS) method are widely used in geodetic science. Most of time, the elements of the coefficient matrix may be consisting of the observations possessing the statistical properties in many applications such as the coordinate transformation (Akyilmaz, 2007; Li et al., 2012; Li et al., 2013; Fang, 2014), and the estimates of the unknown parameters derived by the LS method would not be optimal because the statistical properties of the elements in the coefficient matrix are ignored. The errors-in-variables (EIV) model and so called total least-squares (TLS) method named by Gloub et al. (1980) are more rigorous than the LS method. There are many algorithms to compute the TLS estimate (Gloub et al., 1980; Schaffrin, 2006) or weighted TLS (WTLS) estimate (Schaffrin and Wieser, 2008; Shen et al., 2011; Xu et al., 2012; Amiri-Simkooei and Jazaeri, 2012; Mahboub, 2012; Fang, 2013; Jazaeri et al., 2014).

Unfortunately, like the LS estimate, the WTLS estimate is also extremely vulnerable to the outliers in the EIV model. Although many methods for detecting the outliers in the G-M model are investigated extensively (Baarda, 1968; Pope, 1976; Kok, 1984; Huber 1981; Hekimoglu, 2005; Gui et al. 1999, 2005a, 2005b, 2007, 2011; Guo et al., 2007; Hekimoglu and Erenoglu, 2009; Lehmann, 2013; Hekimoglu et al., 2014), they cannot be directly employed to deal with the outliers in the EIV model. Schaffrin and Uzun (2011) have generalized the mean-shift method to detect a single outlier located either in the observations or in the coefficient matrix in the EIV model. The reliability was also analyzed (Schaffrin and Uzun, 2012). Amiri-Simkooei and Jazaeri (2013) applied the data-snooping procedure to identify the outliers based on the WTLS method formulated with the standard LS theory (Amiri-Simkooei and Jazaeri, 2012). However, the test procedure is required to be implemented more than once while there are some repeated random elements in the different locations of the coefficient matrix like the two-dimensional affine transformation.

The partial EIV model is a generalized EIV model and can avoid considering the correlations between the repeated random elements in the coefficient matrix (Xu et al., 2012). Therefore, it is a more proper model to be used to deal with the case where the coefficient matrix follows a structured characteristic. Unfortunately, the test statistics for detecting the outliers cannot be clearly derived through the existing WTLS method. For this reason, a new two-step iterated approach of computing the WTLS estimates under the framework of LS theory is developed in this paper so that some test statistics of identifying the outliers for the partial EIV model can be constructed.

The remaining of the paper is organized as follows. In Section 2, a two-step iterated method for

the partial EIV model taking advantage of LS theory is proposed. In Section 3, the corresponding w-test statistics are constructed to detect the outliers while the observations, coefficient matrix or both are contaminated with the outliers and an algorithm for detecting outliers in the partial EIV model is designed. If the variance factor is not known, we will employ the least median squares (LMS) method to estimate it. In a latter section, a simulated data and a real data about two-dimensional affine transformation are used to verify the validity of the proposed method. In the end, some concluding remarks are presented.

## 2. PARTIAL EIV MODEL AND WTLS ESTIMATE

As a matter of fact, not all elements of the coefficient matrix are random and there are some repeated random elements in the different locations of the coefficient matrix such as the coordinate transformation. As a result, their correlations between the repeated random elements must be taken into account. The five rules (Mahboub, 2012) can be used to determine the variance-covariance matrix of the coefficient matrix. However, if the partial EIV model proposed by Xu et al. (2012) is considered, the correlations can be avoided so that the additional burden is reduced. Therefore, the partial EIV model is more superior to be adopted. The function model is shown as following:

$$\begin{cases} \mathbf{L} = (\mathbf{X}^T \otimes \mathbf{I}_n)(\mathbf{h} + \mathbf{B}\bar{\mathbf{a}}) + \Delta \\ \mathbf{a} = \bar{\mathbf{a}} + \mathbf{e} \end{cases} \quad (1)$$

Where  $\mathbf{X} = t \times 1$  vector of unknown parameters;  $\mathbf{L} = n \times 1$  vector of observations;  $\mathbf{I}_n = n \times n$  identity matrix;  $\mathbf{h} = nt \times 1$  vector that is consisting of zero and fixed elements of the coefficient matrix  $\mathbf{A}$ ;  $\mathbf{B} = nt \times s$  known structured matrix;  $s$  = the number of different random elements of  $\mathbf{A} = \text{invec}(\mathbf{h} + \mathbf{B}\mathbf{a})$ ;  $\bar{\mathbf{a}} = s \times 1$  true values vector of  $\mathbf{a}$ ;  $\mathbf{e} = s \times 1$  random errors vector of  $\mathbf{a}$ ;  $\Delta = n \times 1$  vector of random errors of observations; *invec* is a mathematic function for transforming an  $nt \times 1$  vector to an  $n \times t$  matrix;  $\otimes$  = Kronecker product operator. The stochastic model is expressed as follows:

$$\begin{bmatrix} \Delta \\ \mathbf{e} \end{bmatrix} \sim N_{n+s} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \mathbf{Q}_L & 0 \\ 0 & \mathbf{Q}_a \end{bmatrix} \right) \quad (2)$$

Where  $\mathbf{Q}_L = n \times n$  cofactor matrix of  $\mathbf{L}$ ;  $\mathbf{Q}_a = s \times s$  cofactor matrix of  $\mathbf{a}$ ;  $\sigma^2$  = unknown variance factor. A two-step iterated method of computing the WTLS estimate for the partial EIV model is proposed in order to develop an outlier detection method suitable for the partial EIV model. For any given  $\mathbf{X}^{(0)}$ , the model (1) can be transformed as follows:

$$\begin{cases} \mathbf{L} - \left[ (\mathbf{X}^{(0)})^T \otimes \mathbf{I}_n \right] \mathbf{h} = \left[ (\mathbf{X}^{(0)})^T \otimes \mathbf{I}_n \right] \mathbf{B}\bar{\mathbf{a}} + \Delta \\ \mathbf{a} = \bar{\mathbf{a}} + \mathbf{e} \end{cases} \quad (3)$$

Furthermore, the model (3) can be rewritten as

$$\begin{bmatrix} \bar{\mathbf{L}} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \left( (\mathbf{X}^{(0)})^T \otimes \mathbf{I}_n \right) \mathbf{B} \\ \mathbf{I}_s \end{bmatrix} \bar{\mathbf{a}} + \begin{bmatrix} \Delta \\ \mathbf{e} \end{bmatrix} \quad (4)$$

where  $\bar{\mathbf{L}} = \mathbf{L} - \left( (\mathbf{X}^{(0)})^T \otimes \mathbf{I}_n \right) \mathbf{h}$ . If we denote

$$\hat{\mathbf{L}} = \begin{bmatrix} \bar{\mathbf{L}} \\ \mathbf{a} \end{bmatrix}, \hat{\mathbf{A}} = \begin{bmatrix} \left( (\mathbf{X}^{(0)})^T \otimes \mathbf{I}_n \right) \mathbf{B} \\ \mathbf{I}_s \end{bmatrix}, cov \begin{bmatrix} \Delta \\ \mathbf{e} \end{bmatrix} = \sigma^2 \mathbf{Q}_{\hat{\mathbf{L}}} = \sigma^2 \begin{bmatrix} \mathbf{Q}_L & 0 \\ 0 & \mathbf{Q}_a \end{bmatrix} \quad (5)$$

the estimate of  $\bar{\mathbf{a}}$  can be derived by the LS principle (Koch, 1999). As a result, we have

$$\hat{\bar{\mathbf{a}}} = \left( \hat{\mathbf{A}}^T \mathbf{Q}_{\hat{\mathbf{L}}}^{-1} \hat{\mathbf{A}} \right)^{-1} \hat{\mathbf{A}}^T \mathbf{Q}_{\hat{\mathbf{L}}}^{-1} \hat{\mathbf{L}} \quad (6)$$

The residual vector of  $\mathbf{a}$  is

$$\mathbf{V}_a = \mathbf{a} - \hat{\bar{\mathbf{a}}} \quad (7)$$

Inserting  $\hat{\bar{\mathbf{a}}}$  into the first equation of the model (1) yields

$$\mathbf{L} = \left( \mathbf{X}^T \otimes \mathbf{I}_n \right) \left( \mathbf{h} + \mathbf{B} \hat{\bar{\mathbf{a}}} \right) + \Delta \quad (8)$$

If the inverse transformation of the mathematic operator *vec* (*invec*) is used, we can obtain

$$\bar{\mathbf{A}} = invvec \left( \mathbf{h} + \mathbf{B} \hat{\bar{\mathbf{a}}} \right) \quad (9)$$

Then the model (8) is easily rewritten as follows:

$$\mathbf{L} = \bar{\mathbf{A}} \mathbf{X} + \Delta \quad (10)$$

Similarly, based on the LS principle (Koch, 1999), the estimate of  $\mathbf{X}$  is

$$\hat{\mathbf{X}} = \left( \bar{\mathbf{A}}^T \mathbf{Q}_L^{-1} \bar{\mathbf{A}} \right)^{-1} \bar{\mathbf{A}}^T \mathbf{Q}_L^{-1} \mathbf{L} \quad (11)$$

and the residual vector of  $\mathbf{L}$  is

$$\mathbf{V}_L = \mathbf{L} - \bar{\mathbf{A}} \hat{\mathbf{X}} \quad (12)$$

The posterior estimate of the variance factor, which can be obtained from Equation 7 and Equation 12, is

$$\hat{\sigma}^2 = \frac{\mathbf{V}_L^T \mathbf{Q}_L^{-1} \mathbf{V}_L + \mathbf{V}_a^T \mathbf{Q}_a^{-1} \mathbf{V}_a}{n - t} \quad (13)$$

### 3. OUTLIER DETECTION PROCEDURE IN PARTIAL EIV MODEL

The data-snooping method suggested by Baarda (1968) is employed extensively in geodetic data processing for detecting the outliers (Kok, 1984; Koch, 1999). If the observations or coefficient matrix in the partial EIV model are contaminated with the outliers, the following  $w$ -test statistics can be constructed based on Equation 6 or Equation 11 to detect the outliers:

$$w_{ai} = \frac{\mathbf{g}_i^T \mathbf{Q}_L^{-1} \mathbf{V}_L}{\sigma \sqrt{\mathbf{g}_i^T \mathbf{Q}_L^{-1} \mathbf{R}_L \mathbf{g}_i}} \sim N(0,1) \quad (14)$$

$$w_{Lj} = \frac{\mathbf{f}_j^T \mathbf{Q}_L^{-1} \mathbf{V}_L}{\sigma \sqrt{\mathbf{f}_j^T \mathbf{Q}_L^{-1} \mathbf{R}_L \mathbf{f}_j}} \sim N(0,1) \quad (15)$$

where  $\mathbf{V}_{\hat{r}} = \hat{\mathbf{L}} - \hat{\mathbf{A}}\hat{\mathbf{a}}$ ,  $\mathbf{R}_{\hat{r}} = \mathbf{I}_{n+s} - \hat{\mathbf{A}}(\hat{\mathbf{A}}^T \mathbf{Q}_L^{-1} \hat{\mathbf{A}})^{-1} \hat{\mathbf{A}}^T \mathbf{Q}_L^{-1}$ ,  $\mathbf{R}_L = \mathbf{I}_n - \bar{\mathbf{A}}(\bar{\mathbf{A}}^T \mathbf{Q}_L^{-1} \bar{\mathbf{A}})^{-1} \bar{\mathbf{A}}^T \mathbf{Q}_L^{-1}$ ;  $\mathbf{g}_i^T = [0, \dots, 1, \dots, 0]_{(n+s) \times 1}$  and  $\mathbf{f}_j^T = [0, \dots, 1, \dots, 0]_{n \times 1}$  are an unit vector with the  $i$ th and  $j$ th element equal to 1, respectively;  $N(0,1)$  represents the standard normal distribution.

In general, when the variance factor is unknown, its posterior estimate  $\hat{\sigma}^2$  can be adopted (Pope, 1976). Then we have

$$\tilde{w}_{ai} = \frac{\mathbf{g}_i^T \mathbf{Q}_L^{-1} \mathbf{V}_L}{\hat{\sigma} \sqrt{\mathbf{g}_i^T \mathbf{Q}_L^{-1} \mathbf{R}_L \mathbf{g}_i}} \sim \tau_n \quad (16)$$

and

$$\tilde{w}_{Lj} = \frac{\mathbf{f}_j^T \mathbf{Q}_L^{-1} \mathbf{V}_L}{\hat{\sigma} \sqrt{\mathbf{f}_j^T \mathbf{Q}_L^{-1} \mathbf{R}_L \mathbf{f}_j}} \sim \tau_{n-t} \quad (17)$$

Where  $\tau_n = \tau$  distribution with  $n$  degree of freedom. The computation about  $\tau$  distribution can be found in Baselga (2007) and Guo and Zhao (2012).

The robust method is an efficient one to estimate the variance factor. By employing the least median squares (LMS) method (Rousseeuw and Leroy, 1987), the variance factor may be estimated by

$$\hat{\sigma}_a = 1.4826 \sqrt{\text{median}(\sigma^2 w_{ai}^2)} \quad (18)$$

or

$$\hat{\sigma}_L = 1.4826 \sqrt{\text{median}(\sigma^2 w_{Lj}^2)} \quad (19)$$

So the test statistics (14) and (15) with (18) and (19) become

$$\bar{w}_{ai} = \frac{\mathbf{g}_i^T \mathbf{Q}_L^{-1} \mathbf{V}}{\hat{\sigma}_a \sqrt{\mathbf{g}_i^T \mathbf{Q}_L^{-1} \mathbf{R} \mathbf{g}_i}} \quad (20)$$

and

$$\bar{w}_{Lj} = \frac{\mathbf{f}_j^T \mathbf{Q}_L^{-1} \mathbf{V}_L}{\hat{\sigma}_L \sqrt{\mathbf{f}_j^T \mathbf{Q}_L^{-1} \mathbf{R}_L \mathbf{f}_j}} \quad (21)$$

The superiority of the above two test statistics is that they are very robust to the outliers so that it is more reliable for them to be used for detecting the outliers. It is to be noted here that they do not strictly follow a normal distribution. Therefore, it is very hard to give the exact probability distributions of them. In order to simplify the computation of the threshold value which is used to identify the outliers, the upper percentage point of the standard normal distribution is still used when the principle of identifying the outliers is established.

The implemented procedure for detecting the outliers in the partial EIV model is summarized as follows:

Step1. Give  $\mathbf{a}, \mathbf{L}, \mathbf{h}, \mathbf{B}, \mathbf{Q}_L, \mathbf{Q}_e$  and define  $\mathbf{Q}_L = \begin{bmatrix} \mathbf{Q}_L & 0 \\ 0 & \mathbf{Q}_e \end{bmatrix}$ .

Step2. Set the initial value  $\hat{\mathbf{X}}^{(0)} = (\mathbf{A}^T \mathbf{Q}_L^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Q}_L^{-1} \mathbf{L}$ .

Step3. For any  $k$ , compute

$$\bar{\mathbf{L}}^{(k)} = \mathbf{L} - \left( (\mathbf{X}^{(k)})^T \otimes \mathbf{I}_n \right) \mathbf{h}, \tilde{\mathbf{L}}^{(k)} = \begin{bmatrix} \bar{\mathbf{L}}^{(k)} \\ \mathbf{a} \end{bmatrix}, \hat{\mathbf{A}}^{(k)} = \begin{bmatrix} \left( (\mathbf{X}^{(k)})^T \otimes \mathbf{I}_n \right) \mathbf{B} \\ \mathbf{I}_t \end{bmatrix}.$$

Step4. Compute  $\hat{\mathbf{a}}^{(k)} = \left( (\hat{\mathbf{A}}^{(k)})^T \mathbf{Q}_L^{-1} \hat{\mathbf{A}}^{(k)} \right)^{-1} (\hat{\mathbf{A}}^{(k)})^T \mathbf{Q}_L^{-1} \tilde{\mathbf{L}}^{(k)}$ .

Step5. Compute  $\bar{\mathbf{A}}^{(k)} = \text{invvec}(\mathbf{h} + \mathbf{B} \hat{\mathbf{a}}^{(k)})$  and  $\hat{\mathbf{X}}^{(k+1)} = \left( (\bar{\mathbf{A}}^{(k)})^T \mathbf{Q}_L^{-1} \bar{\mathbf{A}}^{(k)} \right)^{-1} (\bar{\mathbf{A}}^{(k)})^T \mathbf{Q}_L^{-1} \mathbf{L}$ .

Step6. If  $\|\hat{\mathbf{X}}^{(k+1)} - \hat{\mathbf{X}}^{(k)}\| < \varepsilon$ , the iteration will be stopped, where  $\varepsilon$  is a given value. Otherwise, return to Step 3.

Step7. Compute  $\mathbf{V}_a^{(k)} = \mathbf{a} - \hat{\mathbf{a}}^{(k)}$ ,  $\mathbf{V}_L^{(k)} = \mathbf{L} - \mathbf{A} \hat{\mathbf{X}}^{(k+1)}$  and

$$\hat{\sigma}^2 = \frac{(\mathbf{V}_L^{(k)})^T \mathbf{Q}_L^{-1} \mathbf{V}_L^{(k)} + (\mathbf{V}_a^{(k)})^T \mathbf{Q}_a^{-1} \mathbf{V}_a^{(k)}}{n - t}.$$

Step8. According to the data-snooping procedure, for single outlier, if

$$|\bar{w}_{ai}^{(k)}| > u_{1-\alpha/2} \quad (i \in (n+1, \dots, n+s))$$

and

$$|\bar{w}_{Lj}^{(k)}| > u_{1-\alpha/2} \quad (j \in (1, \dots, n))$$

are satisfied simultaneously, one can judge that the outlier locates in the observation equation containing the observation  $L_j$  and coefficient matrix element  $a_i$ . For multiple outliers, if

$$\max_{i \in \{n+1, \dots, n+s\}} |\bar{w}_{ai}^{(k)}| > u_{1-\alpha/2}$$

and

$$\max_{j \in \{1, \dots, n\}} |\bar{w}_{Lj}^{(k)}| > u_{1-\alpha/2}$$

we will deem that the corresponding observation equation containing the observations  $L_j$  and coefficient matrix elements  $a_i$  is contaminated with outlier. But one still can't confirm that the outliers locate in the observations or coefficient matrix, or both. Here  $u_\alpha$  is the upper  $\alpha$ -percentage point of the standard normal distribution.

Step 9. If multiple outliers exist in the observations or coefficient matrix, the above procedure of Step 1 to Step 8 should be repeated until all the w-test statistics are smaller than the threshold value.

## 4. NUMERICAL RESULTS

### 4.1. Simulated two-dimensional affine transformation

The mathematic model for the two-dimensional affine transformation is expressed as follows:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} x_s & y_s & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_s & y_s & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ c_1 \\ a_2 \\ b_2 \\ c_2 \end{bmatrix} \quad (22)$$

**Table 1:** Coordinates of points with random errors in start system and target system (unit: m)

coordinate	1	2	3	4	5	6	7	8	9	10
$x_s$	70.00	66.16	56.17	43.83	33.82	30.00	33.80	43.83	56.17	66.19
$y_s$	49.98	61.74	69.02	69.01	61.77	50.00	38.25	30.97	30.98	38.24
$x_t$	180.00	141.21	86.70	37.26	11.77	19.99	58.77	113.31	162.77	188.24
$y_t$	59.98	114.67	163.71	188.45	179.38	140.00	85.35	36.28	11.56	20.61

The data are displayed in Table 1, which is taken from Amiri-Simkooei and Jazaeri (2013). In this example, there are ten points in total. So the partial EIV model is

$$\begin{cases} \mathbf{L} = (\mathbf{X}^T \otimes \mathbf{I}_{20})(\mathbf{h} + \mathbf{B}\bar{\mathbf{a}}) + \Delta \\ \mathbf{a} = \bar{\mathbf{a}} + \mathbf{e} \end{cases} \quad (23)$$

where  $\mathbf{a} = [x_s^1, y_s^1, x_s^2, y_s^2 \cdots x_s^{10}, y_s^{10}]^T$ ,  $\mathbf{h}^T = [\mathbf{h}_1^T \mathbf{h}_2^T \cdots \mathbf{h}_6^T]$ ,  $\mathbf{h}_1^T = \mathbf{h}_2^T = \mathbf{h}_4^T = \mathbf{h}_5^T = [0, 0, \dots, 0]_{1 \times 20}$ ,

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \vdots \\ \mathbf{B}_6 \end{pmatrix}, \mathbf{B}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \cdots & 1 & 0 \end{bmatrix}, \mathbf{B}_2 = \begin{bmatrix} 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \cdots & 0 & 1 \end{bmatrix}, \mathbf{B}_3 = \mathbf{B}_6 = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \cdots & 0 & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}_{20 \times 20}$$

$$\mathbf{B}_4 = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}_{20 \times 20}, \mathbf{B}_5 = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}_{20 \times 20}$$

In order to give the reliable evaluations for the proposed outlier detection method, the following five schemes for adding outliers are discussed. The significant level for determining critical value is set as 0.05, which is very frequently used (Gao et al. 1992).

Scheme 1: According to Amiri-Simkooei and Jazaeri (2013), the outlier of magnitude 0.1 m which is 10 times of the priori standard deviation, is added into the xs component of point 4 in the start system.

The residuals of the observations and random vector a and the corresponding w-test statistics are displayed in Table 2. Obviously, the absolute values of residuals of the x components of point 4 in the start system and target system are greater than others. Meanwhile, both  $|\bar{w}_{a27}| = 4.6774$  and

$|\bar{w}_{L7}| = 3.7011$  surpass the threshold value  $u_{0.975} = 1.96$ . So we deem that there is an outlier in the

x component of the start system, target system or both, which is kept the same with the set simulated case. However, we can't determinate the special position of the outlier.

Scheme 2: The outlier of magnitude 0.1 m is added into both components of point 4 in the start system. The residuals and w-test statistics are shown in Table 3. As we know, the absolute values of residuals of the x components of point 4 in both coordinate systems are greater than others. Particularly,

both  $|\bar{w}_{a27}| = 2.7943$  and  $|\bar{w}_{L7}| = 3.2089$  for the x component of point 4 are beyond the threshold

value 1.96, and  $|\bar{w}_{L8}| = 2.8142$  for the yt component of point 4 in the target system exceeds 1.96

too. Although  $|\bar{w}_{a28}| = 1.8708$  for the ys component of point 4 in the start system is smaller than

the threshold value 1.96, the absolute values of w-test statistics and their corresponding absolute values of residuals are very tremendous. Thus, both components of point 4 are considered to be contaminated with outliers. Unfortunately, we can't discriminate the specific positions of these

outliers.

**Table 2:** Residuals of observations and random vector  $\mathbf{a}$  and corresponding  $w$ -test statistics (Scheme 1) (unit: m)

Point No.	Target system			Start system		
	Coordinate	$V_{Lj}$	$\bar{w}_{Lj}$	Coordinate	$V_{ai}$	$\bar{w}_{ai}$
1	$x_t$	-0.0013973	-0.25266	$x_s$	0.0048841	0.32287
	$y_t$	-0.00035329	-0.063881	$y_s$	-0.0013852	-0.091561
2	$x_t$	0.0091772	1.6594	$x_s$	-0.025934	-1.7144
	$y_t$	0.0053935	0.97525	$y_s$	-0.025934	-0.21159
3	$x_t$	0.008246	1.491	$x_s$	-0.029849	-1.9732
	$y_t$	0.0015716	0.28417	$y_s$	0.010229	0.67613
4	$x_t$	<b>-0.020466</b>	<b>-3.7011</b>	$x_s$	<b>0.070755</b>	<b>4.6774</b>
	$y_t$	-0.0055651	-1.0064	$y_s$	-0.018725	-1.2377
5	$x_t$	0.0046209	0.83561	$x_s$	-0.020742	-1.3712
	$y_t$	-0.0011276	-0.20392	$y_s$	0.013769	0.91015
6	$x_t$	0.0036126	0.65332	$x_s$	-0.010178	-0.67285
	$y_t$	0.0021386	0.38674	$y_s$	-0.0013217	-0.087364
7	$x_t$	0.0054174	0.97969	$x_s$	-0.017635	-1.1658
	$y_t$	0.0020203	0.36536	$y_s$	0.0027668	0.18289
8	$x_t$	-0.0040845	-0.7386	$x_s$	0.016878	1.1158
	$y_t$	0.00026835	0.048526	$y_s$	-0.0092558	-0.61182
9	$x_t$	0.00051964	0.093964	$x_s$	-0.0030844	-0.2039
	$y_t$	-0.00050291	-0.090939	$y_s$	0.0030535	0.20184
10	$x_t$	-0.0056462	-1.021	$x_s$	0.014906	0.98537
	$y_t$	-0.0038434	-0.69502	$y_s$	0.0040706	0.26907

**Table 3:** Residuals of observations and random vector  $\mathbf{a}$  and corresponding  $w$ -test statistics (Scheme 2) (unit: m)

Point No.	Target system			Start system		
	Coordinate	$V_{Lj}$	$\bar{w}_{Lj}$	Coordinate	$V_{ai}$	$\bar{w}_{ai}$
1	$x_s$	-0.00093979	-0.10477	$x_s$	0.0052099	0.22502
	$y_t$	0.00072597	0.080929	$y_s$	-0.0047832	-0.20659
2	$x_s$	0.011106	1.238	$x_s$	-0.024541	-1.06
	$y_t$	0.0099526	1.1094	$y_s$	-0.017565	-0.75868
3	$x_s$	0.011362	1.2665	$x_s$	-0.024541	-1.1918
	$y_t$	0.0089374	0.99627	$y_s$	-0.017565	-0.56127
4	$x_s$	<b>-0.028779</b>	<b>-3.2089</b>	$x_s$	<b>0.064696</b>	<b>2.7943</b>
	$y_t$	<b>-0.025239</b>	<b>-2.8142</b>	$y_s$	<b>0.043314</b>	<b>1.8708</b>
5	$x_s$	0.0077303	0.86182	$x_s$	-0.018478	-0.79809
	$y_t$	0.0062289	0.69444	$y_s$	-0.0094334	-0.40745
6	$x_s$	0.0055308	0.61665	$x_s$	-0.0087742	-0.37897
	$y_t$	0.006682	0.745	$y_s$	-0.015647	-0.6758
7	$x_s$	0.0058653	0.65393	$x_s$	-0.017298	-0.7471
	$y_t$	0.0030857	0.34403	$y_s$	-0.00059882	-0.025864
8	$x_s$	-0.0048241	-0.5378	$x_s$	0.016352	0.70624
	$y_t$	-0.0014746	-0.16439	$y_s$	-0.0037588	-0.16235
9	$x_s$	-0.00067131	-0.074838	$x_s$	-0.0039448	-0.17038
	$y_t$	-0.0033185	-0.36994	$y_s$	0.011924	0.51502
10	$x_s$	-0.0063798	-0.71127	$x_s$	0.014371	0.62072
	$y_t$	-0.0055804	-0.62215	$y_s$	0.0095432	0.41219

Scheme 3: The outlier of magnitude 0.1 m is added into the  $x_s$  component of point 4 in the start system and the  $y_t$  component of point 4 in target system.

The residuals and the  $w$ -test statistics are obtained, which is displayed in Table 4. The results from Table 4 show that the test statistics satisfy  $|\bar{w}_{a27}| = 4.9415 > 1.96$  and  $|\bar{w}_{L7}| = 4.106 > 1.96$ , which shows that the  $x$  component of point 4 is possibly contaminated with an outlier. Although the absolute value of residual for the  $y_t$  component of point 4 in the target system is small,  $|\bar{w}_{a28}| = 3.0386 > 1.96$  and the absolute value of residual for the  $y_s$  component of point 4 in the start system demonstrate that there is an outlier in the  $y$  component.

**Table 4:** Residuals of observations and random vector  $\mathbf{a}$  and corresponding  $w$ -test statistics (Scheme3) (unit: m)

Point No.	Target system			Start system		
	Coordinate	$V_{Lj}$	$\bar{w}_{Lj}$	Coordinate	$V_{ai}$	$\bar{w}_{ai}$
1	$x_t$	-0.0017266	-0.49177	$x_s$	0.0053326	0.42223
	$y_t$	-0.00078794	-0.22442	$y_s$	-0.00030473	-0.024126
2	$x_t$	0.0077803	2.2158	$x_s$	-0.024022	-1.9021
	$y_t$	0.0035539	1.0121	$y_s$	0.0013597	0.10765
3	$x_t$	0.0059859	1.7047	$x_s$	-0.02675	-2.1181
	$y_t$	-0.0014011	-0.399	$y_s$	0.017599	1.3933
4	$x_t$	<b>-0.014411</b>	<b>-4.106</b>	$x_s$	<b>0.062409</b>	<b>4.9415</b>
	$y_t$	<b>0.0023763</b>	<b>0.67706</b>	$y_s$	<b>-0.038379</b>	<b>-3.0386</b>
5	$x_t$	0.002359	0.67191	$x_s$	-0.01763	-1.396
	$y_t$	-0.0040971	-1.167	$y_s$	0.021125	1.6725
6	$x_t$	0.0022133	0.63048	$x_s$	-0.0082465	-0.65296
	$y_t$	0.00030441	0.086714	$y_s$	0.0032151	0.25455
7	$x_t$	0.0050847	1.4484	$x_s$	-0.017166	-1.3592
	$y_t$	0.001589	0.45262	$y_s$	0.0038252	0.30285
8	$x_t$	-0.0035534	-1.0121	$x_s$	0.01616	1.2795
	$y_t$	0.00097174	0.27677	$y_s$	-0.011008	-0.8715
9	$x_t$	0.0013812	0.39339	$x_s$	-0.004261	-0.33738
	$y_t$	0.00063268	0.1802	$y_s$	0.00023425	0.018546
10	$x_t$	-0.0051132	0.39339	$x_s$	0.014175	1.1224
	$y_t$	-0.0031418	0.1802	$y_s$	0.0023335	0.18475

**Table 5:** Residuals of observations and random vector  $\mathbf{a}$  and corresponding  $w$ -test statistics (Scheme 4) (unit: m)

Point No.	Target system			Start system		
	Coordinate	$V_{Lj}$	$\bar{w}_{Lj}$	Coordinate	$V_{ai}$	$\bar{w}_{ai}$
1	$x_t$	-0.0023489	-0.58721	$x_s$	0.0090658	0.75905
	$y_t$	-0.00016403	-0.041006	$y_s$	-0.0040362	-0.33795
2	$x_t$	0.0051497	1.2873	$x_s$	-0.0082485	-0.69062
	$y_t$	0.00618	1.5449	$y_s$	-0.014416	-1.207
3	$x_t$	0.0017338	0.4334	$x_s$	-0.0012538	-0.10498
	$y_t$	0.0028432	0.71071	$y_s$	-0.0079015	-0.6616
4	$x_t$	-0.0030607	-0.76547	$x_s$	-0.0056531	-0.47332
	$y_t$	<b>-0.008957</b>	<b>-2.2401</b>	$y_s$	<b>0.029688</b>	<b>2.4858</b>
5	$x_t$	-0.0018874	-0.47184	$x_s$	0.0078336	0.65589
	$y_t$	0.00014316	0.035789	$y_s$	-0.0043421	-0.36357
6	$x_t$	-0.00040699	-0.10176	$x_s$	0.0074691	0.62537
	$y_t$	0.0029242	0.73114	$y_s$	-0.012501	-1.0467
7	$x_t$	0.0044704	1.1177	$x_s$	-0.013485	-1.1291
	$y_t$	0.0021985	0.54968	$y_s$	0.00014176	0.01187
8	$x_t$	-0.0025448	-0.63621	$x_s$	0.010115	0.84692
	$y_t$	-3.0876e-005	-0.0077191	$y_s$	-0.0049597	-0.41528
9	$x_t$	0.0030058	0.75143	$x_s$	-0.014008	-1.1728
	$y_t$	-0.00099484	-0.2487	$y_s$	0.0099803	0.83567
10	$x_t$	-0.0041109	-1.0278	$x_s$	0.0081649	0.68362
	$y_t$	-0.0041423	-1.0356	$y_s$	0.0083459	0.69881

**Table 6:** Residuals of observations and random vector  $a$  and corresponding  $w$ -test statistics with deleting point 4 in both of start system and target system (unit: m)

Point No.	Target system			Start system		
	Coordinate	$V_{Lj}$	$\bar{w}_{Lj}$	Coordinate	$V_{ai}$	$\bar{w}_{ai}$
1	$x_t$	-0.0025167	-0.7071	$x_s$	0.0087558	0.89974
	$y_t$	-0.00065539	-0.18414	$y_s$	-0.0024088	-0.24753
2	$x_t$	0.0044403	1.2804	$x_s$	-0.0095588	-0.98226
	$y_t$	0.0041034	1.1833	$y_s$	-0.0075354	-0.77434
3	$x_t$	0.0005872	0.17764	$x_s$	-0.0033716	-0.34646
	$y_t$	-0.00051201	-0.15489	$y_s$	0.0032211	0.331
4	$x_t$	—	—	$x_s$	—	—
	$y_t$	—	—	$y_s$	—	—
5	$x_t$	-0.0030325	-0.91736	$x_s$	0.0057185	0.58763
	$y_t$	-0.0032077	-0.97033	$y_s$	0.0067668	0.69537
6	$x_t$	-0.0011139	-0.3212	$x_s$	0.0061635	0.63336
	$y_t$	0.00085512	0.24659	$y_s$	-0.0056457	-0.58015
7	$x_t$	0.0043048	1.2096	$x_s$	-0.013791	-1.4172
	$y_t$	0.0017142	0.48168	$y_s$	0.001748	0.17963
8	$x_t$	-0.002273	-0.64022	$x_s$	0.010617	1.091
	$y_t$	0.00076394	0.21517	$y_s$	-0.0075976	-0.78073
9	$x_t$	0.0034443	0.97631	$x_s$	-0.013198	-1.3562
	$y_t$	0.00028881	0.081865	$y_s$	0.0057284	0.58866
10	$x_t$	-0.0038405	-1.0817	$x_s$	0.0086644	0.89036
	$y_t$	-0.0033504	-0.94369	$y_s$	0.0057231	0.58811

**Table 7:** Residuals of observations and random vector  $a$  and corresponding  $w$ -test statistics (Scheme 5) (unit: m)

Point No.	Target system			Start system		
	Coordinate	$V_{Lj}$	$\bar{w}_{Lj}$	Coordinate	$V_{ai}$	$\bar{w}_{ai}$
1	$x_t$	0.0045553	0.82093	$x_s$	-0.014464	-0.88826
	$y_t$	0.0018727	0.33749	$y_s$	0.0016235	0.09968
2	$x_t$	<b>-0.016465</b>	<b>-2.9676</b>	$x_s$	<b>0.052397</b>	<b>3.2177</b>
	$y_t$	-0.0067103	-1.2095	$y_s$	-0.0061015	-0.37462
3	$x_t$	0.0059838	1.0783	$x_s$	-0.026717	-1.6407
	$y_t$	-0.0014032	-0.25286	$y_s$	0.017593	1.0802
4	$x_t$	0.0098111	1.7687	$x_s$	-0.013972	-0.85803
	$y_t$	<b>0.012636</b>	<b>2.2779</b>	$y_s$	<b>-0.03093</b>	<b>-1.8991</b>
5	$x_t$	-0.0039266	-0.70769	$x_s$	0.0021913	0.13456
	$y_t$	-0.0067598	-1.2183	$y_s$	0.019193	1.1784
6	$x_t$	-0.0040698	-0.73355	$x_s$	0.011557	0.70971
	$y_t$	-0.002357	-0.42483	$y_s$	0.0012864	0.078982
7	$x_t$	0.0012096	0.21801	$x_s$	-0.004942	-0.30348
	$y_t$	-5.41e-005	-0.0097511	$y_s$	0.0026377	0.16195
8	$x_t$	-0.0035487	-0.63958	$x_s$	0.016128	0.9904
	$y_t$	0.00097413	0.17556	$y_s$	-0.011002	-0.67549
9	$x_t$	0.0052762	0.9509	$x_s$	-0.016531	-1.0151
	$y_t$	0.0022808	0.41105	$y_s$	0.0014335	0.088012
10	$x_t$	0.0011737	0.21154	$x_s$	-0.0056471	-0.34679
	$y_t$	-0.00047889	-0.086313	$y_s$	0.0042659	0.26192

Scheme 4: The outlier of magnitude 0.1m is added into the  $y$  component of point 4 in both start system and target system.

The concrete results are presented in Table 5. It is not difficult to know  $|\bar{w}_{a28}| = 2.4858 > 1.96$  and  $|\bar{w}_{L8}| = 2.2401 > 1.96$  from Table 5, but the absolute values of other  $w$ -statistics are smaller than 1.96. It means that only  $y$  component of point 4 contains an outlier, which is consistent with the set simulated case. If we will delete point 4 in both coordinate systems, the new results about the residuals and  $w$ -test statistics are obtained, which is displayed in Table 6. It is shown that all  $|\bar{w}_{ai}|$  and  $|\bar{w}_{Lj}|$  are smaller than the threshold value 1.96, which demonstrates that the remaining observations are clean without the effects of outliers.

We just discuss the case that the outlier locates in the same point in two different systems for scheme 1 to 4. In fact, there may be multiple outliers in the different points for the two-dimensional coordinate transformation. Hence, the following scheme 5 is used to assess the efficiency of the proposed procedure for detecting multiple outliers in the partial EIV model.

Scheme 5: In this simulation, two outliers of magnitude 0.1 m are added to the  $x_s$  component of point 2 in the start system and the  $y_t$  component of point 4 in the target system, respectively.

The detail results about the residuals and  $w$ -test statistics are listed in Table 7.  $|\bar{w}_{a23}| = 3.2177 > 1.96$  and  $|\bar{w}_{L3}| = 2.9676 > 1.96$  indicate that the  $x$  component of point 2 contains an outlier. On the other hand, due to  $|\bar{w}_{a28}| = 1.8991$  and  $|\bar{w}_{L8}| = 2.2779 > 1.96$ , the  $y$  component of point 4 is probable to be contaminated with an outlier. Because the outlier may locate in the different locations, we will delete point 2 in both coordinate systems firstly. After that, the new results and  $w$ -test statistics are obtained, which can be found in Table 8. Apparently, there is an outlier in  $y$  component in the start system or target system or both based on the criterion for identifying outlier. As a result, point 4 in both two coordinate systems should be deleted. After removing the assigned outlying observations, the new results about the residuals and  $w$ -test statistics are presented in Table 9, which indicates that there is no outlier in the observations of both coordinate systems.

**Table 8:** Residuals of observations and random vector  $\mathbf{a}$  and corresponding  $w$ -test statistics with deleting the point 2 in both of start system and target system (Scheme 5) (unit: m)

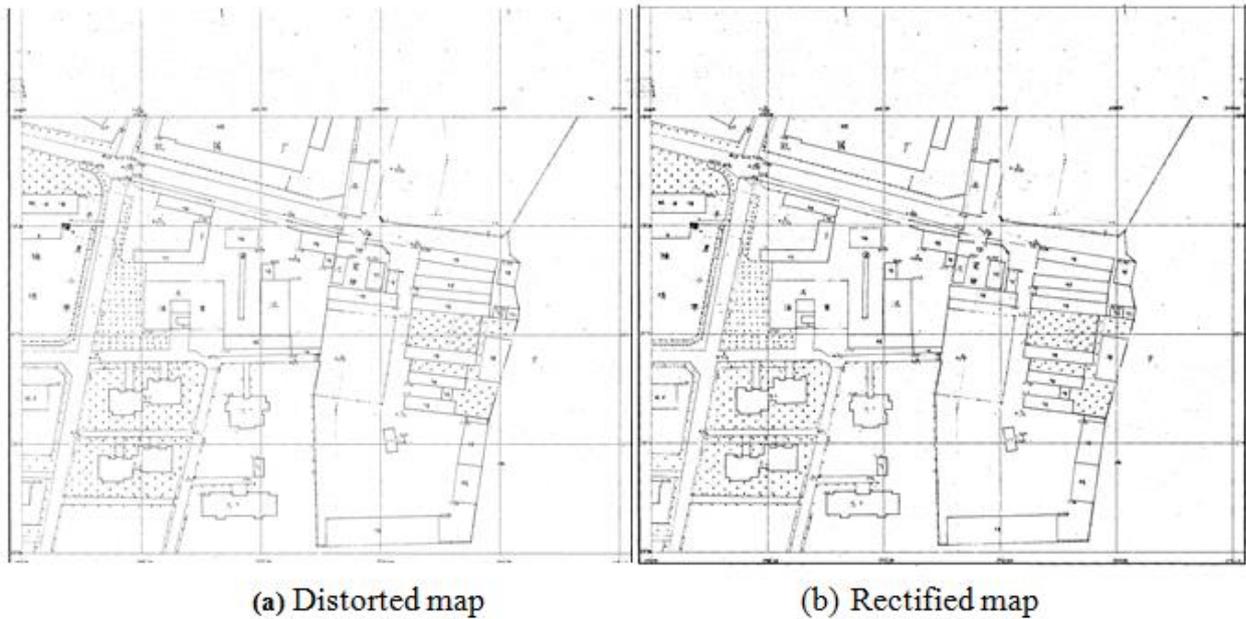
Point No.	Target system			Start system		
	Coordinate	$V_{Lj}$	$\bar{w}_{Lj}$	Coordinate	$V_{ai}$	$\bar{w}_{ai}$
1	$x_t$	-0.0015997	-0.48961	$x_s$	0.0051277	0.51616
	$y_t$	-0.00063521	-0.19442	$y_s$	-0.00065729	-0.06616
2	$x_t$	—	—	$x_s$	—	—
	$y_t$	—	—	$y_s$	—	—
3	$x_t$	-0.00017461	-0.053438	$x_s$	-0.0071247	-0.71718
	$y_t$	-0.003914	-1.1979	$y_s$	0.015311	1.5411
4	$x_t$	0.005999	1.7512	$x_s$	-0.0018417	-0.18538
	$y_t$	<b>0.011082</b>	<b>3.2349</b>	$y_s$	<b>-0.032343</b>	<b>-3.2555</b>
5	$x_t$	-0.0048223	-1.3711	$x_s$	0.005046	0.50794
	$y_t$	-0.0071243	-2.0256	$y_s$	0.018861	1.8985
6	$x_t$	-0.0026122	-0.74458	$x_s$	0.0069248	0.69706
	$y_t$	-0.0017621	-0.50227	$y_s$	0.0018267	0.18387
7	$x_t$	0.0035602	1.0213	$x_s$	-0.012434	-1.2516
	$y_t$	0.00090246	0.25889	$y_s$	0.0035085	0.35315
8	$x_t$	-0.0020919	-0.59622	$x_s$	0.011502	1.1578
	$y_t$	0.0015692	0.44725	$y_s$	-0.010462	-1.053
9	$x_t$	0.0043731	1.2433	$x_s$	-0.013669	-1.3759
	$y_t$	0.0019111	0.54335	$y_s$	0.0010985	0.11057
10	$x_t$	-0.0026315	-0.76792	$x_s$	0.0064686	0.65114
	$y_t$	-0.002029	-0.59209	$y_s$	0.0028559	0.28746

**Table 9:** Residuals of observations and random vector  $\mathbf{a}$  and corresponding  $w$ -test statistics with deleting point 2 and point 4 in both of start system and target system (Scheme 5) (unit: m)

Point No.	Target system			Start system		
	Coordinate	$V_{Lj}$	$\bar{w}_{Lj}$	Coordinate	$V_{ai}$	$\bar{w}_{ai}$
1	$x_t$	-0.00070383	-0.23467	$x_s$	0.004853	0.58449
	$y_t$	0.00102	0.34007	$y_s$	-0.0054855	-0.66066
2	$x_t$	—	—	$x_s$	—	—
	$y_t$	—	—	$y_s$	—	—
3	$x_t$	0.0027479	1.0525	$x_s$	-0.0080217	-0.96613
	$y_t$	0.0014848	0.56868	$y_s$	-0.00044471	-0.053561
4	$x_t$	—	—	$x_s$	—	—
	$y_t$	—	—	$y_s$	—	—
5	$x_t$	-0.002371	-0.79065	$x_s$	0.0042931	0.51706
	$y_t$	-0.0025964	-0.8658	$y_s$	0.0056432	0.67966
6	$x_t$	-0.0012774	-0.40134	$x_s$	0.0065153	0.7847
	$y_t$	0.00070397	0.22118	$y_s$	-0.0053684	-0.64656
7	$x_t$	0.0036943	1.1405	$x_s$	-0.012475	-1.5025
	$y_t$	0.0011501	0.35506	$y_s$	0.0027853	0.33546
8	$x_t$	-0.0027834	-0.85898	$x_s$	0.011715	1.4109
	$y_t$	0.00029229	0.090205	$y_s$	-0.0067323	-0.81083
9	$x_t$	0.0035459	1.0946	$x_s$	-0.013415	-1.6157
	$y_t$	0.00038276	0.11816	$y_s$	0.005557	0.66929
10	$x_t$	-0.0028525	-0.8962	$x_s$	0.0065361	0.7872
	$y_t$	-0.0024375	-0.76582	$y_s$	0.0040453	0.48721

## 4.2 Real data about map rectification

The example is about the map rectification. The 2D affine transformation is used to rectify the map. The scale of map is 1:500 for figure 1.



**Figure 1:** The distorted map and its rectified map using affine transformation

There are ten common points whose theoretical coordinates are previously known, and then we sample their coordinates on the distorted map. The affine transformation is used to rectify the map. The sampled coordinates and theoretical coordinates of common points are treated as the coordinates in the start system and target system, respectively, which is displayed in Table 10. The transformation parameters can be estimated by using the common points with the 2D affine transformation. By employing the proposed algorithm, the residuals and  $w$ -test statistics of the observations and random vector  $\mathbf{a}$  are derived, which is shown in Table 11. Because the  $w$ -test statistics satisfy  $|\bar{w}_{L14}| = 21.838 > 1.96$  and  $|\bar{w}_{a34}| = 21.172 > 1.96$ , the point 7 is suspected as an outlier and should be deleted. Then the new residuals and  $w$ -test statistics are obtained, which can be found in Table 12. Due to  $|\bar{w}_{a35}| = 1.7622 < 1.96$  for point 9 in the target system, there are no outliers in the observations even if  $|\bar{w}_{L15}| = 2.297 > 1.96$  according to the criterion for identifying the outliers in section 3. Therefore, the only outlier is identified. After that, the transformation parameters are estimated by the WTLS method. The results are presented in Table 13. By checking the reliability of the proposed method, the fifteen non-common points are employed to evaluate the performance of the proposed algorithm and RMSE (Root mean square error) is used to judge the influence of outlier for the coordinates. The RMSE for the data-snooping procedure is 0.00892, but is 0.032786 for the WTLS method with outliers. The reason is that the transformation parameters estimated by the WTLS method are disturbed with the outliers.

**Table 10:** Coordinates of ten common points and fifteen non-common points in both coordinate systems (unit: cm)

Common point				Non-common point			
Start system		Target system		Start system		Target system	
$u_s$	$v_s$	$u_t$	$v_t$	$u_s$	$v_s$	$u_t$	$v_t$
77.58677125	87.246990015	34.0	85.0	28.17098162	87.272316176	19.0	85.0
28.13210239	103.72201572	19.0	90.0	44.65580551	87.245720424	24.0	85.0
77.58934311	103.71908529	34.0	90.0	61.10492273	87.244663745	29.0	85.0
28.12765661	120.18027351	19.0	95.0	94.04265529	87.236868084	39.0	85.0
77.606496	120.160256564	34.0	95.0	110.52943807	87.236555391	44.0	85.0
77.61204959	136.623478492	34.0	100	44.62795576	103.714775203	24.0	90.0
28.10320572	154.068706679	19.0	105	61.0884887	103.713279354	29.0	90.0
77.62038088	153.103856739	34.0	105	94.05693187	103.715676958	39.0	90.0
28.08255946	169.529298616	19.0	110.0	110.52749417	103.705978224	44.0	90.0
77.59748129	169.545714888	34.0	110.0	44.64684942	120.159237522	24.0	95.0
				61.11220165	120.128950724	29.0	95.0
				94.0829795	120.1725203	39.0	95.0
				110.5677384	120.165699479	44.0	95.0
				28.11643765	136.617556801	19.0	100.0
				44.60626576	136.611885876	24.0	100.0

**Table 11:** Residuals of observation and random vector  $a$  and corresponding  $w$ -test statistics for mapping rectification (unit: cm)

Point No.	Target system			Start system		
	Coordinate	$V_{Lj}$	$\bar{w}_{Lj}$	Coordinate	$V_{ai}$	$\bar{w}_{ai}$
1	$x_t$	0.0054641	0.66522	$x_s$	-0.0016239	-0.58441
	$y_t$	-0.02307	-2.8086	$y_s$	0.006993	2.5164
2	$x_t$	-0.0048054	-0.58676	$x_s$	0.0013928	0.50126
	$y_t$	0.045573	5.5647	$y_s$	-0.013814	-4.971
3	$x_t$	0.0042668	0.4667	$x_s$	-0.0012697	-0.45693
	$y_t$	-0.016875	-1.8457	$y_s$	0.0051151	1.8406
4	$x_t$	-0.0040557	-0.45558	$x_s$	0.0011516	0.41444
	$y_t$	0.055618	6.2476	$y_s$	-0.016859	-6.0667
5	$x_t$	-0.00098049	-0.10244	$x_s$	0.00030013	0.10801
	$y_t$	-0.0021115	-0.22059	$y_s$	0.00064008	0.23033
6	$x_t$	-0.0030061	-0.31407	$x_s$	0.000902	0.32462
	$y_t$	0.0065435	0.68363	$y_s$	-0.0019835	-0.71373
7	$x_t$	0.0018414	0.20717	$x_s$	-0.00028704	-0.1033
	$y_t$	<b>-0.1941</b>	<b>-21.838</b>	$y_s$	<b>0.058838</b>	<b>21.172</b>
8	$x_t$	-0.0058015	-0.63472	$x_s$	0.0017438	0.62757
	$y_t$	0.010432	1.1414	$y_s$	-0.0031622	-1.1379
9	$x_t$	0.0070075	0.85308	$x_s$	-0.0022537	-0.81107
	$y_t$	0.092938	11.314	$y_s$	-0.028173	-10.138
10	$x_t$	6.9481e-005	0.0084591	$x_s$	-5.6045e-005	-0.02017
	$y_t$	0.025052	3.0499	$y_s$	-0.007594	-2.7326

**Table 12:** Residuals of observation and random vector  $\mathbf{a}$  and corresponding  $w$ -test statistics for mapping rectification with deleting point 7 (unit: cm)

Point No.	Target system			Start system		
	Coordinate	$V_{Lj}$	$\bar{w}_{Lj}$	Coordinate	$V_{ai}$	$\bar{w}_{ai}$
1	$x_t$	0.005221	1.3414	$x_s$	-0.0015825	-1.1719
	$y_t$	0.0026575	0.68277	$y_s$	-0.00080753	-0.59671
2	$x_t$	-0.0043552	-1.1521	$x_s$	0.0013201	0.97753
	$y_t$	-0.0019646	-0.51971	$y_s$	0.00059699	0.44113
3	$x_t$	0.0041199	0.94397	$x_s$	-0.0012487	0.97753
	$y_t$	-0.0014459	-0.33129	$y_s$	0.00043918	0.44113
4	$x_t$	-0.0035079	-0.86199	$x_s$	0.0010633	0.78736
	$y_t$	-0.0022192	-0.54533	$y_s$	0.00067433	0.49828
5	$x_t$	-0.0010283	-0.22444	$x_s$	0.00031166	0.23079
	$y_t$	0.0030519	0.6661	$y_s$	-0.00092718	-0.68512
6	$x_t$	-0.002957	-0.64539	$x_s$	0.00089626	0.66369
	$y_t$	0.0014165	0.30916	$y_s$	-0.00043027	-0.31794
7	$x_t$	—	—	$x_s$	—	—
	$y_t$	—	—	$y_s$	—	—
8	$x_t$	-0.0056571	-1.2965	$x_s$	0.0017147	1.2698
	$y_t$	-0.0049908	-1.1438	$y_s$	0.0015164	1.1205
9	$x_t$	0.0078511	2.297	$x_s$	-0.0023797	-1.7622
	$y_t$	0.0041787	1.2226	$y_s$	-0.0012697	-0.93825
10	$x_t$	0.00031337	0.080512	$x_s$	-9.4978e-005	-0.070332
	$y_t$	-0.00068396	-0.17573	$y_s$	0.00020779	0.15354

**Table 13:** Transformation parameters estimated by the WTLS method before deleting outlier and after deleting outlier (unit: cm)

Before deleting outlier	After deleting outlier
0.30309255593699	0.30310519134397
0.00003187394065	0.00002566590120
10.4752902610926	10.47510689386349
0.00139656637130	0.00000654387860
0.30313281644081	0.30381576309241
58.46940628440629	58.48957855017623

## 5. CONCLUSIONS

The WTLS estimate of the partial EIV model may strongly be influenced by the outliers. The aim of this paper is to develop an approach to detect the outliers in the partial EIV model. Firstly, we propose a two-step iterated method of computing the WTLS estimates for the partial EIV model based on the standard LS theory. Then the corresponding  $w$ -test statistics are constructed to detect the outliers while the observations, coefficient matrix or both are contaminated with the outliers. If the variance factor is unknown, it may be estimated by the LMS method. Making using of the proposed two-step iterated method, the implement algorithm for detecting the outliers in the partial EIV model is proposed. Through the numerical results with the

two-dimensional affine transformation, the identification of outliers is implemented only once through the proposed procedure compared with previously approach while single outlier is considered. For multiple outliers, the repeated test with step by step is suggested. However, we still can't discriminate that the outliers locate in the observation or coefficient matrix or both, which is a very open problem to be discussed in the future.

## ACKNOWLEDGEMENTS

The authors are grateful to the editor and two anonymous reviewers for their constructive comments and suggestions so that the paper is substantial improved. This research has been supported jointly by the National Natural Science Foundation of China (Grant No. 41174005, 414714009) and State Key Laboratory of Geodesy and Earth's Dynamics (SKLGED 2017-3-2-E)

## REFERENCES

- Akyilmaz, O. "Total least squares solution of coordinate transformation." *Survey Review* 39 (303) (2007): 68-80.
- Amiri-Simkooei, A.R. and Jazaeri, S. "Data-snooping procedure applied to errors-in-variable models." *Stud.Geophys.Geod.* 57 (2013): 426-441.
- Amiri-Simkooei, A.R. and Jazaeri, S. "Weighted total least-squares formulated by standard least-squares theory." *Journal of Geodetic Science* 2(2) (2012): 113-124.
- Baarda, W. "A testing procedure for use in geodetic networks." Publications on Geodesy 2(5) (1968), Netherlands Geodetic Commission, Delft, the Newtherlands.
- Baselga, S. "Critical limitation in use of  $\tau$  test for gross error detection." *Jounral of Surveying Engineering* 133(2) (2007): 52-55.
- Fang, X. "A total least squares solution for geodetic datum transformations." *Acta Geod Geophys* 49 (2014): 189-207.
- Fang, X. "Weighted total least squares: necessary and sufficient conditions, fixed and random parameters." *Journal of Geodesy* 87 (2013): 733-749.
- Gao, Y., Krakiwsky, E.J. and Czompo J. "Robust testing procedure for detection of multiple blunders." *Journal of Surveying Engineering* 118 (1992): 11-23.
- Gloub, G.H. and Van Loan, C. F. "An analysis of the total least squares problem." *SIAM J. Numer. Anal.* 17(6) (1980): 883-893.
- Gui, Q.M. and Liu, J.S. "Generalized shrunken-type robust estimation." *Jounral of Surveying Engineering* 125(4) (1999):177-184.
- Gui, Q.M., Gong, Y.S., Li, G.Z. and Li, B.L. "A Bayesian approach to the detection of gross errors based on posterior probability." *Journal of Geodesy* 81(2007): 651-659.
- Gui, Q.M., Li G.Z., and Ou J.K. "Robust-biased estimation based on quasi-accurate detection." *Jounral of Surveying Engineering* 131(3) (2005a): 67-72.

- Gui, Q.M., Li G.Z., and Ou J.K. "Robust-Type Biased Estimation in Gauss-Markov Model." *Survey Review* 38(298) (2005b): 299-307.
- Gui, Q.M., Li, X., Gong, Y.S., Li, B.L. and Li, G.Z. "Bayesian unmasking method for locating multiple gross errors based on posterior probabilities of classification variables." *Journal of Geodesy* 85 (2011): 191-203.
- Guo, J.F. and Zhao, J. "Comparative analysis of statistical tests used for detection and identification of outliers (in Chinese)." *Acta Geodaetica et Cartographica Sinica* 41(1) (2012): 14-18.
- Guo, J.F., Ou, J.K. and Wang, H.T. "Quasi-accurate detection of outliers for correlated observations." *Journal of Surveying Engineering* 133 (2007): 129-133.
- Hekimoglu, S. and Erenoglu, R.C. "New method for outlier diagnostics in linear regression." *Journal of Surveying Engineering* 135(3) (2009): 85-89.
- Hekimoglu, S. "Do robust methods identify outliers more reliably than conventional tests for outliers?" *ZFV* 130 (2005): 174-180.
- Hekimoglu, S., Erdogan, B., Soyacan, M. and Durdag, U.M. "Univariate approach for detecting outliers in geodetic networks." *Journal of Surveying Engineering* 140 (2014): 1-8.
- Huber, P.J. *Robust Statistics*. New York: John Wiley & Sons, 1981.
- Jazaeri, S., Amiri-Simkooei, A.R. and Sharifi, M.A. "Iterative algorithm for weighted total least squares adjustment." *Survey Review* 46(334) (2014): 19-27.
- Koch, K.R. "*Parameter Estimation and Hypothesis Testing in Linear Models*, 2nd Edition." Berlin: Springer-Verlag, Berlin, 1999.
- Kok, J.J. "*On data snooping and multiple outliers testing*." Rochville: NOAA Technical Report, Nos NGS 30, 1984.
- Lehmann, R. "On the formulation of the alternative hypothesis for geodetic outlier detection." *Journal of Geodesy* 87 (2013): 373-386.
- Li, B.F., Shen Y.Z., Zhang X.F., Li C. and Lou L.Z. "Seamless multivariate affine error-in-variables transformation and its application to map rectification." *International Journal of Geographical Information Science* 27(8) (2012): 1572-1592.
- Li, B.F., Shen, Y.Z. and Lou, L.Z. "Noniterative datum transformation revisited with two-dimensional affine model as a case study." *Journal of Surveying Engineering* 139 (2013): 166-175.
- Mahboub, V. "On weighted total least-squares for geodetic transformations." *Journal of Geodesy* 86 (2012): 359-367.
- Pope, A.J. "*The statistics of residuals and the detection of outliers*." Rochville: NOAA Technical report, Nos 65, NGS 1, 1976.
- Rousseeuw, P.J. and Leroy, A.M. "*Robust Regression and Outliers Detection*." New York: John Wiley & Sons, 1987.
- Schaffrin, B. "A note on constrained total least-squares estimation." *Linear Algebra and its Applications* 417 (2006): 245-258.

Schaffrin, B. and Uzun, S. “Errors-in-variables for mobile algorithms in the presence of outliers.” *Archives of Photogrammetry, Cartography and Remote Sensing* 22 (2011): 377-387.

Schaffrin, B. and Uzun, S. On the reliability of errors-in-variables models. *Acta Et Commentationes Universitatis Tartuensis De Mathematica* 16(1) (2012): 69-81.

Schaffrin, B. and Wieser, A. “On weighted total least-squares adjustment for linear regression.” *Journal of Geodesy* 82 (2008): 415-421.

Shen, Y.Z., Li, B.F. and Chen, Y. “An iterative solution of weighted total least-squares adjustment.” *Journal of Geodesy* 85 (2011): 229-238.

Xu, P.L., Liu, J.N. and Shi, C. “Total least-squares adjustment in partial errors-in-variables models: algorithm and statistical analysis.” *Journal of Geodesy* 86 (2012): 661-675.

Recebido em 18 de junho de 2015.

Aceito em 11 de junho de 2016.