



An extended w -test for outlier diagnostics in linear models

Yangkang Yu¹ · Ling Yang¹ · Yunzhong Shen¹

Received: 9 November 2023 / Accepted: 22 April 2024 / Published online: 18 June 2024
© Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

The issue of outliers has been a research focus in the field of geodesy. Based on a statistical testing method known as the w -test, data snooping along with its iterative form, iterative data snooping (IDS), is commonly used to diagnose outliers in linear models. However, in the case of multiple outliers, it may suffer from the masking and swamping effects, thereby limiting the detection and identification capabilities. This contribution is to investigate the cause of masking and swamping effects and propose a new method to mitigate these phenomena. First, based on the data division, an extended form of the w -test with its reliability measure is presented, and a theoretical reinterpretation of data snooping and IDS is provided. Then, to alleviate the effects of masking and swamping, a new outlier diagnostic method and its iterative form are proposed, namely data refining and iterative data refining (IDR). In general, if the total observations are initially divided into an inlying set and an outlying set, data snooping can be considered a process of selecting outliers from the inlying set to the outlying set. Conversely, data refining is then a reverse process to transfer inliers from the outlying set to the inlying one. Both theoretical analysis and practical examples show that IDR would keep stronger robustness than IDS due to the alleviation of masking and swamping effect, although it may pose a higher risk of precision loss when dealing with insufficient data.

Keywords Outliers · w -test · Data snooping · Iterative data snooping (IDS) · Masking and swamping effects · Extended w -test · Data refining · Iterative data refining (IDR)

1 Introduction

In many geodetic data processing tasks, large sets of observations are recorded or sampled, where it is nearly impossible that such datasets are free from outliers (Lehmann 2013; Rofatto et al. 2020a). In case outliers occur, the least squares (LS) estimation, typically applied to geodetic data processing, may lose the properties of unbiasedness and minimum variance on the parameter estimate (Koch 1999; Teunissen 2000). Thus, one of the major challenges of geodetic data analysis is to deal with outliers properly. In this paper, the outlier follows the definition given by Lehmann (2013): outliers are observations that are contaminated by gross errors. In contrast, the observations without any contamination are defined as inliers (Hawkins 1980). Generally, there have been two main categories of methods developed to deal with outliers, which are outlier diagnostics and robust estimation (Rousseeuw and Leroy 1987).

The methods of outlier diagnostics aim to pinpoint outliers from the data, after which these outliers are to be removed or corrected, followed by an LS analysis for the remaining cases. The outlier diagnostics has a long history, since Thompson (1935), Pearson and Sekar (1936), Nair (1948), and Grubbs (1969) studied the outliers based on the normalized or studentized residuals of LS. Their results were followed up by many researchers who detect and identify outliers in normal samples or linear regressions, e.g., Daniel (1960), Anscombe (1960), Quesenberry and David (1961), Ferguson (1961), Srikantan (1961), David and Paulson (1965), Stefansky (1972), Ellenberg (1973, 1976), Rosner (1975), Galpin and Hawkins (1981), and Fischler and Bolles (1981). Besides, there are extensive reviews of the subject by Beckman and Cook (1983).

In geodesy, considering the characteristics of geodetic data, outlier diagnostics methods have also been extensively studied. After over half a century of development, data snooping has become one of the best-established outlier diagnostics methods and has been used as a standard procedure for quality control. Data snooping with its associated reliability measure originated in the pioneering work of (Baarda 1967, 1968),

✉ Ling Yang
lingyang@tongji.edu.cn; ling_yangsy@163.com

¹ Department of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China

later extended by Pope (1976) to the case that the precision of the observations is unknown, see also, e.g., Alberda (1976), Kok (1984), Teunissen (1985), Xu (1987a, b) and Koch (1999). Data snooping usually consists of three steps. First, the test statistics are usually used to detect the presence of outliers in the observation system (Teunissen 2000; Koch 2015). Then, model selection is carried out to identify the most possible outlier by screening each observation (Förstner 1983; Yang et al. 2013). Finally, parameter estimation is carried out after excluding the tested outlier. The statistical test used in data snooping is well-known as the w -test. In case that observations are independent of each other, the w -test statistics are equivalent to the normalized or studentized least squares residuals.

For modern geodetic applications, there are typically large datasets that are very likely to contain multiple outliers. In this case, one of the most commonly used methods is to implement the data snooping procedure iteratively, processing outliers one by one (Mickey et al. 1967; Barnett and Lewis 1978; Gentle 1978), which is known as iterative data snooping (IDS) (Kok 1984; Lehmann and Scheffler 2011; Rofatto et al. 2017; Klein et al. 2022). However, this approach is not theoretically rigorous, since it is assumed that there is only one outlier in each iteration, but this assumption is overturned immediately in the next iteration (Lehmann and Lösler 2016). Actually, in the case of multiple outliers, the testing methods would be hampered by the masking and swamping effects. Specifically, masking effects are that multiple outliers can mask each other easily, which increases the difficulty of outlier detection (McMillan 1971). In addition, swamping effects are that if the suspected maximum number of outliers is large, the statistical test tends to declare more outliers than there are (Fieller 1976).

Another kind of method, robust estimation, aims to attain a solution with higher robustness via modifying the score function of LS. Since Box (1953) coined the term ‘robustness’, an enormous amount of studies has been published on this subject, e.g., M-estimation (Huber 1964; Hampel et al. 1986), L-estimation (Sarhan and Greenberg 1956), R-estimation (Hodges and Lehmann 1963; Jaeckel 1972; Duchnowski 2013), S-estimation (Rousseeuw and Yohai 1984), median estimation (Stigler 1977; Duchnowski 2010), quantile regression (Koenker and Hallock 2001), Msplit estimation (Wiśniewski 2009, 2010), least absolute values method (Edgeworth 1887; Khodabandeh and Amiri-Simkooei 2011), least median of squares method (Rousseeuw 1984; Rousseeuw and Leroy 1987), sign-constrained robust least squares (Xu 2005), and various generalized versions of them. Using the theory of breakdown point (Hodges 1967; Donoho and Huber 1983), they have achieved satisfactory performance when dealing with multiple outliers.

Among all of these robust estimations, the subclass of M-estimation is more attractive in geodesy, since they are

computationally efficient and can be easily implemented in existing geodetic adjustment software (Yang et al. 2002; Koch 2013). Compared to other robust estimations, M-estimation is usually consistent with the LS solution in the absence of outliers, so that retains higher precision of parameter estimates in most cases. Due to the nonlinear property of M-estimators, the iteratively reweighted least squares (IRLS) procedure is developed for practical applications (Holland and Welsch 1977; Koch 1999). The principle of IRLS is to adapt the weight for each observation based on the previous adjustment residuals. Such weights reduce or eliminate the effect of outliers on the final estimate of the parameters via many different down-weight strategies, e.g., Huber’s estimator (Huber 1964), Danish method (Krarup et al. 1980), Hampel’s estimator (Hampel et al. 1986), IGGIII method (Yang 1994; Yang et al. 2002). However, since these methods are based on the measure of initial LS residuals, the masking and swamping effects unavoidably bring in some incorrect down-weight selection (Hekimoğlu 1997, 1999).

The present contribution is to investigate the cause of masking and swamping effects and to propose a new method to mitigate these phenomena. First, based on the data division, an extended form of the w -test with its reliability measure is presented, and a theoretical reinterpretation of data snooping and IDS is provided. Then, a new outlier diagnostic method and its iterative form are proposed, namely data refining and iterative data refining (IDR). In general, if the total observations are initially divided into an inlying set and an outlying set, data snooping can be considered a process of selecting outliers from the inlying set to the outlying set. Conversely, data refining is then a reverse process to transfer inliers from the outlying set to the inlying one. For IDS, all data are usually assumed as inliers in the initial stage. In this case, the inlying set is probably contaminated when there are multiple outliers. Consequently, a contaminated inlying set might invalidate the test decision, which might cause the masking and swamping effect. However, in the initial stage of IDR, the suspected outliers are moved out of the inlying set. Therefore, a reliable inlying set relatively guarantees the validity of the following test, thereby effectively alleviating the masking and swamping effect.

This contribution is structured as follows: In Sect. 2, the Gauss–Markov model is briefly reviewed. Also, Baarda’s w -test and data snooping are introduced. In Sect. 3, first, based on the data division, the extended w -test with its associated reliability measure is presented. Then, data snooping is reinterpreted, and a new method called data refining is proposed. In Sect. 4, the iterative forms of data snooping and data refining are presented, which are IDS and IDR, respectively. As well, the comparison of IDS and IDR is discussed in detail. In Sect. 5, a linear fitting example is used to analyze the property of IDR for dealing with outliers. Finally, Sect. 6 concludes this contribution and discusses future work.

2 Baarda’s data snooping in the linear models

In this section, the linear model with least squares estimates is briefly reviewed. Then, the w -test and data snooping with its associated reliability measure are introduced.

2.1 Linear models and least squares (LS) estimate

Consider the linear model of observation equations (Koch 1999; Teunissen 2000):

$$E(\mathbf{y}) = \mathbf{A}\mathbf{x}, \quad D(\mathbf{y}) = \sigma^2 \mathbf{Q}. \tag{1}$$

Here, $E(\cdot)$ and $D(\cdot)$ are the expectation and dispersion operators, respectively. $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the non-random design matrix of $\text{rank}(\mathbf{A}) = n < m$. $\mathbf{y} \in \mathbb{R}^m$ is the vector of observations with normal distribution, and $\mathbf{x} \in \mathbb{R}^n$ is the vector of unknown parameters. The symmetric positive matrix $\mathbf{Q} \in \mathbb{R}^{m \times m}$ denotes the variance–covariance cofactor of \mathbf{y} , and σ^2 denotes the variance of unit weight. The LS estimate of the unknown parameters \mathbf{x} is given by:

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{y}, \tag{2}$$

with the cofactor matrix of variance–covariance:

$$\mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} = (\mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A})^{-1}. \tag{3}$$

The residual vector of \mathbf{y} can be given as:

$$\hat{\mathbf{e}} = \left[\mathbf{I}_m - \mathbf{A}(\mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Q}^{-1} \right] \mathbf{y}, \tag{4}$$

with the cofactor matrix of variance–covariance:

$$\mathbf{Q}_{\hat{\mathbf{e}}\hat{\mathbf{e}}} = \mathbf{Q} - \mathbf{A}(\mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A})^{-1} \mathbf{A}^T. \tag{5}$$

For condition equations, the linear model in Eq. (1) can be equivalently written as (Koch 1999; Teunissen 2000):

$$\mathbf{B}^T E(\mathbf{y}) = 0, \quad D(\mathbf{y}) = \sigma^2 \mathbf{Q}. \tag{6}$$

Here, $\mathbf{B} \in \mathbb{R}^{m \times (m-n)}$ is a basis matrix satisfied that $\mathbf{B}^T \mathbf{A} = 0$ and $\text{rank}(\mathbf{B}) = m - n$. In this case, the residual vector can be given as:

$$\hat{\mathbf{e}} = \mathbf{Q}\mathbf{B}(\mathbf{B}^T \mathbf{Q}\mathbf{B})^{-1} \mathbf{B}^T \mathbf{y}, \tag{7}$$

with the cofactor matrix of variance–covariance:

$$\mathbf{Q}_{\hat{\mathbf{e}}\hat{\mathbf{e}}} = \mathbf{Q}\mathbf{B}(\mathbf{B}^T \mathbf{Q}\mathbf{B})^{-1} \mathbf{B}^T \mathbf{Q}. \tag{8}$$

Particularly, if σ^2 is unknown, it can be estimated as:

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{e}}^T \mathbf{Q}^{-1} \hat{\mathbf{e}}}{m - n}. \tag{9}$$

In general, LS is the best linear unbiased estimator (BLUE) for linear models (Koch 1999; Teunissen 2000). However, the optimal properties of LS might be compromised, once the observations are contaminated from gross errors resulting in the presence of outliers (Rousseeuw and Leroy 1987; Lehmann 2013). Therefore, statistical test procedures for outlier diagnostics have been developed.

2.2 w -test and data snooping

Suppose there is a suspected gross error in the k th observation, then the linear model in Eqs. (1) or (6) becomes (Koch 1999; Teunissen 2000):

$$\begin{cases} E(\mathbf{y}) = \mathbf{A}\mathbf{x} + \mathbf{c}_k \nabla_k, & D(\mathbf{y}) = \sigma^2 \mathbf{Q} \\ \mathbf{B}^T E(\mathbf{y}) = \mathbf{B}^T \mathbf{c}_k \nabla_k, & D(\mathbf{y}) = \sigma^2 \mathbf{Q} \end{cases}, \quad k \in \{1, \dots, m\}, \tag{10}$$

where ∇_k is the size of the gross error in the k th observation. $\mathbf{c}_k = [0, \dots, 0, 1, 0, \dots, 0]^T$ is a unit vector with the k th element equal to one. The w -test statistic for the k th observation can then be formed as follows (Baarda 1967; Teunissen 2000):

$$w_k = \frac{\hat{\nabla}_k}{\sigma \sqrt{q_{\hat{\nabla}_k}^2}} = \frac{\mathbf{c}_k^T \mathbf{M} \mathbf{y}}{\sigma \sqrt{\mathbf{c}_k^T \mathbf{M} \mathbf{c}_k}}, \tag{11}$$

with

$$\begin{aligned} \mathbf{M} &= \mathbf{Q}^{-1} - \mathbf{Q}^{-1} \mathbf{A} (\mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Q}^{-1} \\ &= \mathbf{B} (\mathbf{B}^T \mathbf{Q}\mathbf{B})^{-1} \mathbf{B}^T = \mathbf{Q}^{-1} \mathbf{Q}_{\hat{\mathbf{e}}\hat{\mathbf{e}}} \mathbf{Q}^{-1}. \end{aligned} \tag{12}$$

Here, $\hat{\nabla}_k = \mathbf{c}_k^T \mathbf{M} \mathbf{y} / \mathbf{c}_k^T \mathbf{M} \mathbf{c}_k$ is the LS estimate of ∇_k , and $q_{\hat{\nabla}_k}^2 = 1 / \mathbf{c}_k^T \mathbf{M} \mathbf{c}_k$ is the variance cofactor of $\hat{\nabla}_k$.

In specific, w_k obeys a normal distribution, that is $w_k \sim N(\delta, 1)$, where $\delta = \frac{\nabla_k}{\sigma \sqrt{q_{\hat{\nabla}_k}^2}}$. If there is no gross error in observation y_k , then $\delta = 0$, otherwise, $\delta \neq 0$. Therefore, one can test whether the data y_k is an inlier or outlier according to the significance of w_k . The so-called w -test is organized as follows. Given a critical value k_α , if $|w_k| > k_\alpha$, then y_k is tested as an inlier; otherwise, y_k is tested as an outlier. Here, k_α is calculated as the quantile of $N(0, 1)$ upon a significance level α , that is $k_\alpha = N_{1-\alpha/2}(0, 1)$.

Note that when σ in Eq. (11) is unknown, then it can be estimated as:

$$\hat{\sigma}_k^2 = \frac{\hat{\mathbf{e}}_k^T \mathbf{Q}^{-1} \hat{\mathbf{e}}_k}{m - n - 1} = \frac{\mathbf{y}^T [\mathbf{M} - \mathbf{M} \mathbf{c}_k (\mathbf{c}_k^T \mathbf{M} \mathbf{c}_k)^{-1} \mathbf{c}_k^T \mathbf{M}] \mathbf{y}}{m - n - 1}, \tag{13}$$

where $\hat{\mathbf{e}}_k$ is the observation residuals calculated by Eq. (10). In this case, $w_k = \frac{\hat{v}_k}{\hat{\sigma}_k \sqrt{q_{\hat{v}_k}^2}}$ turns to obey a student distribution

with a degree of freedom $m - n - 1$ and a noncentralized parameter δ , that is $w_k \sim t(m - n - 1, \delta)$ (Xu 1987a, b). Correspondingly, the critical value of the w -test becomes $k_\alpha = t_{1-\alpha/2}(m - n - 1, 0)$. Also, w_k can be equivalently transformed as another test statistic obeying the τ distribution with a degree of freedom 1 and $m - n - 1$, that is $\frac{\sqrt{m-n}w_k}{\sqrt{m-n-1+w_k^2}} \sim \tau(1, m - n - 1, \delta)$ (Pope 1976; Koch 1999; Lehmann 2012).

Furthermore, considering the reliability measure, the minimal detectable bias (MDB) can be given by (Baarda 1967, 1968; Teunissen 2000):

$$\text{MDB}_k = \delta_0 \sigma \sqrt{q_{\hat{v}_k}^2} = \frac{\delta_0 \sigma}{\sqrt{\mathbf{c}_k^T \mathbf{M} \mathbf{c}_k}}. \tag{14}$$

Here, δ_0 is the theoretical noncentralized parameter, that can be computed via $\delta_0 = N_{1-\frac{\alpha}{2}}(0, 1) - N_\beta(0, 1)$ or $\delta_0 = t_{1-\frac{\alpha}{2}}(m - n - 1, 0) - t_\beta(m - n - 1, 0)$, where α and β are the significance level and the power of the test, respectively.

By letting k run from 1 to and including m , one can screen the whole data set for potential outliers. The significance of the test statistics is tested by comparing them to the critical values. Thus, the observation with the largest absolute value of test statistic is tested to be an outlier when:

$$\max_{k \in \{1, \dots, m\}} |w_k| > k_\alpha. \tag{15}$$

The procedure for screening each observation for an outlier is known as ‘‘data snooping’’ (Kok 1984; Teunissen 2000). Furthermore, in the case of multiple outliers, the data snooping procedure can be implemented iteratively to process outliers one by one, which is known as iterative data snooping (IDS) (Kok 1984; Lehmann and Scheffler 2011; Rofatto et al. 2017; Klein et al. 2022).

Note that if observations are independent of each other, the w -test statistics are equivalent to the normalized or studentized least squares residuals used by Thompson (1935),

Pearson and Sekar (1936), Nair (1948), and Grubbs (1969), as:

$$w_k = \frac{\hat{e}_k}{\sigma \sqrt{q_{\hat{e}_k}^2}} \text{ or } \frac{\hat{e}_k}{\hat{\sigma}_k \sqrt{q_{\hat{e}_k}^2}}, \tag{16}$$

where \hat{e}_k is the k th element of the LS residuals $\hat{\mathbf{e}}$, and $q_{\hat{e}_k}^2$ are the variance cofactor of \hat{e}_k .

3 Data division and extended w -test

In this section, to distinguish inliers and outliers, the data division is first discussed. Based on the data division, an extended w -test with its associated reliability is presented. Then, data snooping is reinterpreted, and a new method called data refining is proposed.

3.1 Data division

By measuring via a suitably standardized scale, observations can be divided into two groups based on the characteristics of inlying and outlying, referred to as inliers and outliers (Hawkins 1980; Xu 1987a, 1987b). In other words, let all data come from a complete set $\{y_1, \dots, y_m\}$, then they can always be divided into two sets, called the inlying set and the outlying set. Generally, for an observation system, the observations in the inlying set are to be retained, while those in the outlying set are to be excluded.

Here, a data division method is proposed as follows. Given an outlier number q , then all of the candidate pairs of the inlying-outlying set can be listed as follows:

$$I_i \cup O_i = \{y_1, \dots, y_m\}, \quad i \in \left\{ 1, \dots, \binom{m}{q} \right\}. \tag{17}$$

Each pair of the inlying and outlying sets corresponds to a candidate model (Teunissen 2000, 2018; Lehmann and Lösler 2016):

$$\left\{ \begin{array}{l} \mathbf{E}(\mathbf{y}) = \mathbf{A} \mathbf{x} + \mathbf{C}_i \mathbf{b}_i, \quad \mathbf{D}(\mathbf{y}) = \sigma^2 \mathbf{Q} \\ \mathbf{B}^T \mathbf{E}(\mathbf{y}) = \mathbf{B}^T \mathbf{C}_i \mathbf{b}_i, \quad \mathbf{D}(\mathbf{y}) = \sigma^2 \mathbf{Q} \end{array} \right., \quad i \in \left\{ 1, \dots, \binom{m}{q} \right\}, \tag{18}$$

where $\mathbf{C}_i \in \mathbb{R}^{m \times q}$ is a design matrix which consists of unit vectors generated by the outlying set O_i , and $\mathbf{b}_i \in \mathbb{R}^q$ is the size vector of gross errors in the data of the outlying set. Based on the principle of model selection, one can find the most likely pair of inlying-outlying sets, I and O , with the least square of residuals as:

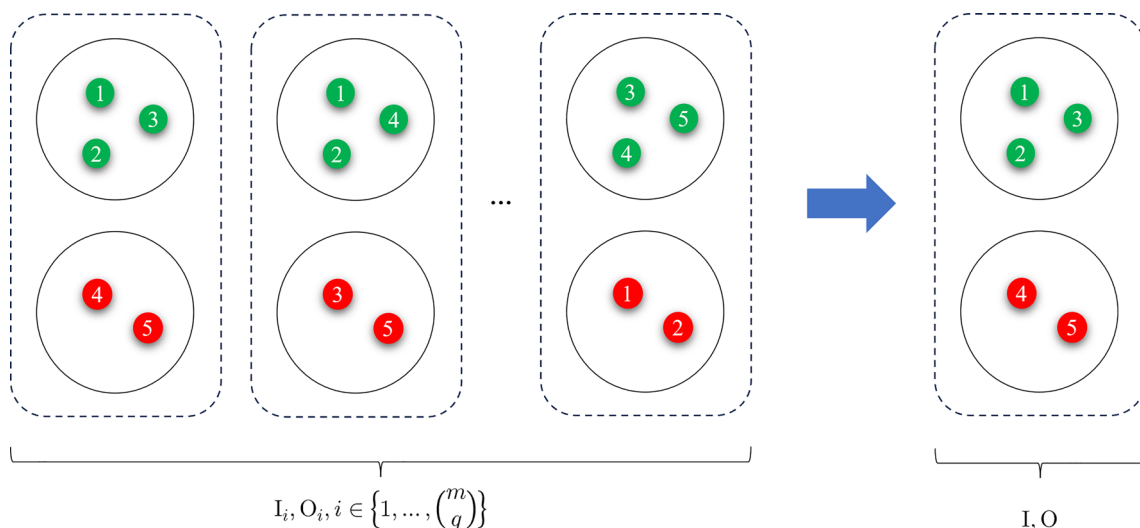


Fig. 1 Procedure of data division

$$\begin{aligned}
 C &= \underset{C_i, i \in \left\{1, \dots, \binom{m}{q}\right\}}{\operatorname{argmin}} \left\{ \hat{e}_i^T Q^{-1} \hat{e}_i \right\} \\
 &= \underset{C_i, i \in \left\{1, \dots, \binom{m}{q}\right\}}{\operatorname{argmin}} \left\{ y^T \left[M - MC_i (C_i^T MC_i)^{-1} C_i^T M \right] y \right\},
 \end{aligned}
 \tag{19}$$

where \hat{e}_i is the observation residuals calculated via Eq. (18). Note that, when $q = 0$, there is only one candidate pair of inlying-outlying set, that is $I_i = \{y_1, \dots, y_m\}$, and $O_i = \emptyset$. In addition, when $q = m - n$, \hat{e}_i always equals the zero vector and the data division will be invalidated. Therefore, the choice of q should satisfy that $0 \leq q \leq m - n - 1$. Figure 1 gives an example of data division where $m = 5$ and $q = 2$.

3.2 Extended w -test

After the data division, the statistical test can then be used to test data in both the inlying set and the outlying set. This procedure usually consists of following two phases. The initial phase is shown in Fig. 2. Using a data division method such as Eq. (19) with a presumed *suspected outlier number* q_0 , the total data can be initially divided into two subsets: an initial inlying set denoted as I_0 and an initial outlying set denoted as O_0 . The data in I_0 and O_0 are thus considered the *suspected inliers* and the *suspected outliers*, respectively. In this case, the linear model is given by:

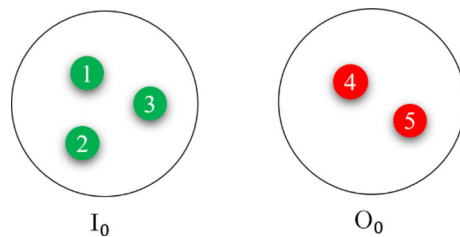


Fig. 2 Initial phase of the test

$$\begin{cases} E(y) = Ax + C_0 b_0, & D(y) = \sigma^2 Q \\ B^T E(y) = B^T C_0 b_0, & D(y) = \sigma^2 Q \end{cases},
 \tag{20}$$

where $C_0 \in \mathbb{R}^{m \times q_0}$ is a design matrix generated by O_0 , and $b_0 \in \mathbb{R}^{q_0}$ is the gross error size vector of data in O_0 . The model given in Eq. (20) is called the *initial model*. Considering the specific structure of C_0 , this model is essentially equivalent to Eqs. (1) and (6) with the data in O_0 excluded.

In the testing phase, a statistical test can be applied to determine whether a suspected data, originating from either the initial inlying or outlying set, should be classified as an outlier or an inlier. As shown in Fig. 3, if the suspected data y_k is selected, all data can be divided into three disjoint parts, the inlying set I_k , the outlying set O_k , and the testing set $\{y_k\}$, in which the data number are $m - 1 - q_k$, q_k , and 1, respectively. In this case, the linear model in Eq. (20) becomes:

$$\begin{cases} E(y) = Ax + C_k b_k + c_k \nabla_k, & D(y) = \sigma^2 Q \\ B^T E(y) = B^T C_k b_k + B^T c_k \nabla_k, & D(y) = \sigma^2 Q \end{cases},
 \tag{21}$$

$k \in \{1, \dots, m\}$,

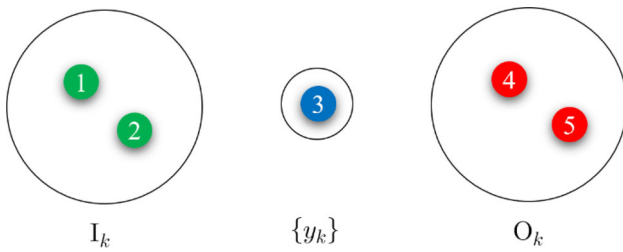


Fig. 3 Testing phase of the test ($k = 3$)

where $\mathbf{c}_k \in \mathbb{R}^m$ is the unit design vector generated by $\{y_k\}$, $\mathbf{C}_k \in \mathbb{R}^{m \times q_k}$ is the design matrix generated by \mathbf{O}_k . Correspondingly, $\mathbf{b}_k \in \mathbb{R}^{q_k}$ denotes the gross error sizes of data in \mathbf{O}_k , and ∇_k represents the gross error size of y_k . This model is called the testing model, which is equivalent to Eq. (10) with the data in \mathbf{O}_k excluded.

Using the model given in Eq. (21), the LS estimate of ∇_k is derived as:

$$\hat{\nabla}_k = \frac{\mathbf{c}_k^T [\mathbf{M} - \mathbf{M}\mathbf{C}_k(\mathbf{C}_k^T\mathbf{M}\mathbf{C}_k)^{-1}\mathbf{C}_k^T\mathbf{M}]\mathbf{y}}{\mathbf{c}_k^T [\mathbf{M} - \mathbf{M}\mathbf{C}_k(\mathbf{C}_k^T\mathbf{M}\mathbf{C}_k)^{-1}\mathbf{C}_k^T\mathbf{M}]\mathbf{c}_k} \tag{22}$$

Proof See Appendix.

The variance cofactor of $\hat{\nabla}_k$ is given by:

$$q_{\hat{\nabla}_k}^2 = \frac{1}{\mathbf{c}_k^T [\mathbf{M} - \mathbf{M}\mathbf{C}_k(\mathbf{C}_k^T\mathbf{M}\mathbf{C}_k)^{-1}\mathbf{C}_k^T\mathbf{M}]\mathbf{c}_k} \tag{23}$$

Proof See Appendix.

Then, according to Eqs. (22) and (23), the w -test statistics can be extended as:

$$w_k = \frac{\hat{\nabla}_k}{\sigma\sqrt{q_{\hat{\nabla}_k}^2}} = \frac{\mathbf{c}_k^T [\mathbf{M} - \mathbf{M}\mathbf{C}_k(\mathbf{C}_k^T\mathbf{M}\mathbf{C}_k)^{-1}\mathbf{C}_k^T\mathbf{M}]\mathbf{y}}{\sigma\sqrt{\mathbf{c}_k^T [\mathbf{M} - \mathbf{M}\mathbf{C}_k(\mathbf{C}_k^T\mathbf{M}\mathbf{C}_k)^{-1}\mathbf{C}_k^T\mathbf{M}]\mathbf{c}_k}} \tag{24}$$

Likewise, the extended w -test statistic w_k satisfies that $w_k \sim N(\delta, 1)$, where $\delta = \frac{\nabla_k}{\sigma\sqrt{q_{\hat{\nabla}_k}^2}}$. If y_k is an inlier, then $\delta = 0$; otherwise, $\delta \neq 0$. Therefore, following the principle of significance test (Fisher 1925), the extended w -test can be organized as follows. Given a critical value $k_\alpha = N_{1-\frac{\alpha}{2}}(0, 1)$, if $|w_k| \leq k_\alpha$, then y_k is tested as an inlier; otherwise, y_k is tested as an outlier.

In addition, if σ in Eq. (24) is unknown, it can be estimated via Eq. (21), as:

$$\hat{\sigma}_k^2 = \frac{\mathbf{y}^T [\mathbf{M} - \mathbf{M}\mathbf{G}_k(\mathbf{G}_k^T\mathbf{M}\mathbf{G}_k)^{-1}\mathbf{G}_k^T\mathbf{M}]\mathbf{y}}{m - n - q_k - 1} \tag{25}$$

Proof See Appendix.

Here, $\mathbf{G}_k \in \mathbb{R}^{m \times (q_k+1)}$ is the design matrix generated by $\mathbf{O}_k \cup \{y_k\}$, that is $\mathbf{G}_k = [\mathbf{C}_k \ \mathbf{c}_k]$. Correspondingly, the test statistic and the critical value become $w_k = \frac{\hat{\nabla}_k}{\hat{\sigma}_k\sqrt{q_{\hat{\nabla}_k}^2}} \sim t(m - n - q_k - 1, \delta)$ and $k_\alpha = t_{1-\alpha/2}(m - n - q_k - 1, 0)$. Also, w_k can be equivalently transformed as another test statistic, that is $\frac{\sqrt{m-n-q_k}w_k}{\sqrt{m-n-q_k-1+w_k^2}} \sim \tau(1, m - n - q_k - 1, \delta)$.

Furthermore, considering the reliability measure, the MDB of the extended w -test can be given as:

$$\begin{aligned} \text{MDB}_k &= \delta_0\sigma\sqrt{q_{\hat{\nabla}_k}^2} \\ &= \frac{\delta_0\sigma}{\sqrt{\mathbf{c}_k^T [\mathbf{M} - \mathbf{M}\mathbf{C}_k(\mathbf{C}_k^T\mathbf{M}\mathbf{C}_k)^{-1}\mathbf{C}_k^T\mathbf{M}]\mathbf{c}_k}} \end{aligned} \tag{26}$$

where δ_0 is computed by $N_{1-\frac{\alpha}{2}}(0, 1) - N_\beta(0, 1)$ or $t_{1-\frac{\alpha}{2}}(m - n - q_k - 1, 0) - t_\beta(m - n - q_k - 1, 0)$.

Note that in the w -test for IDS or other iterative testing methods, the test statistic in Eq. (24) has been used as an equivalent form of the classical one in Eq. (11) to test data within the observation system, i.e., in the inlying set \mathbf{I}_0 (Kok 1984; Teunissen 1990, 2000). Additionally, in the extended w -test, both the format and the usage of this test statistic are further extended to encompass a broader range of the testing data, specifically for data outside the observation system, i.e., in the outlying set \mathbf{O}_0 . For example, using the extended w -test, one can first choose either data in \mathbf{I}_0 with the largest $|w_k|$ or that in \mathbf{O}_0 with the smallest $|w_k|$ as the most suspected data. Then, this suspected data can be tested as an outlier or an inlier via evaluating the significance of the test statistic.

Essentially, the extended w -test is based on the principle of using the inlying set to test whether the testing data is an outlier. Consequently, the test performance depends on the number and quality of data in the inlying set. In specific, from Eq. (26), one can see that a larger sample size for the inlying set results in a reduced MDB, thus enhancing the test power. Additionally, apart from the substantial capacity, maintaining the purity of the inlying set is also crucial, since the contamination of the inlying set might cause the masking or swamping effects. For instance, if there are outliers left in the inlying set \mathbf{I}_k during the testing phase, the linear model in Eq. (21) will become:

$$\begin{cases} \mathbf{E}(\mathbf{y}) = \mathbf{A}\mathbf{x} + \mathbf{C}_k\mathbf{b}_k + \mathbf{c}_k\nabla_k + \mathbf{C}_l\mathbf{b}_l, \mathbf{D}(\mathbf{y}) = \sigma^2\mathbf{Q} \\ \mathbf{B}^T\mathbf{E}(\mathbf{y}) = \mathbf{B}^T\mathbf{C}_k\mathbf{b}_k + \mathbf{B}^T\mathbf{c}_k\nabla_k + \mathbf{B}^T\mathbf{C}_l\mathbf{b}_l, \mathbf{D}(\mathbf{y}) = \sigma^2\mathbf{Q} \end{cases} \tag{27}$$

where $C_l \in \mathbb{R}^{m \times q_l}$ is the design matrix generated by the outliers left in I_k , and $b_l \in \mathbb{R}^{q_l}$ denotes the corresponding gross error sizes of these outliers. In this case, the estimate $\hat{\nabla}_k$ in Eq. (22) will be biased:

$$E(\hat{\nabla}_k) = \frac{c_k^T [M - MC_k (C_k^T MC_k)^{-1} C_k^T M] E(y)}{c_k^T [M - MC_k (C_k^T MC_k)^{-1} C_k^T M] c_k} = \nabla_k + \frac{c_k^T [M - MC_k (C_k^T MC_k)^{-1} C_k^T M] C_l b_l}{c_k^T [M - MC_k (C_k^T MC_k)^{-1} C_k^T M] c_k} \tag{28}$$

One can see that the expectation of $w_k = \frac{\hat{\nabla}_k}{\sigma \sqrt{q_{\hat{\nabla}_k}^2}}$ will not be zero even if $\nabla_k = 0$. Consequently, some inliers would be identified as outliers, which can be called the swamping effect. Moreover, if the observation precision is unknown, the estimate of σ^2 in Eq. (25) will also be biased:

$$E(\hat{\sigma}_k^2) = \frac{E\{y^T [M - MC_k (C_k^T MC_k)^{-1} C_k^T M] y\}}{m - n - q_k - 1} = \sigma^2 + \frac{b_l^T C_l^T [M - MC_k (C_k^T MC_k)^{-1} C_k^T M] C_l b_l}{m - n - q_k - 1} \tag{29}$$

Here, since $\frac{b_l^T C_l^T [M - MC_k (C_k^T MC_k)^{-1} C_k^T M] C_l b_l}{m - n - q_k - 1} > 0$, this estimate $\hat{\sigma}_k^2$ will be enlarged by the outliers left in I_k , thereby shrinking the size of $w_k = \frac{\hat{\nabla}_k}{\hat{\sigma}_k \sqrt{q_{\hat{\nabla}_k}^2}}$. In this case, the outliers can be difficult to detect, which is exactly the masking effect.

3.3 Data snooping

Using the extended w -test, data snooping can be reinterpreted as follows. Assuming there is at most one outlier in the initial inlying set I_0 , data snooping is a procedure of traversing all data in the inlying set to find out this outlier.

The procedure of data snooping is given in Fig. 4. Specifically, according to Eqs. (24) and (25), one can construct a test statistic w_k for each data in the initial inlying set, $y_k \in I_0$. In this case, for each w_k , the inlying set I_k , outlying set O_k , and testing set $\{y_k\}$ are constructed as follows:

$$I_k \cup \{y_k\} = I_0 \text{ and } O_k = O_0. \tag{30}$$

Then, the data with the largest test statistic is considered the most suspected outlier in I_0 . Generally, data snooping consists of the following three parts.

Detection: with a significance level α , the significance of the largest test statistic is tested by comparing it with the

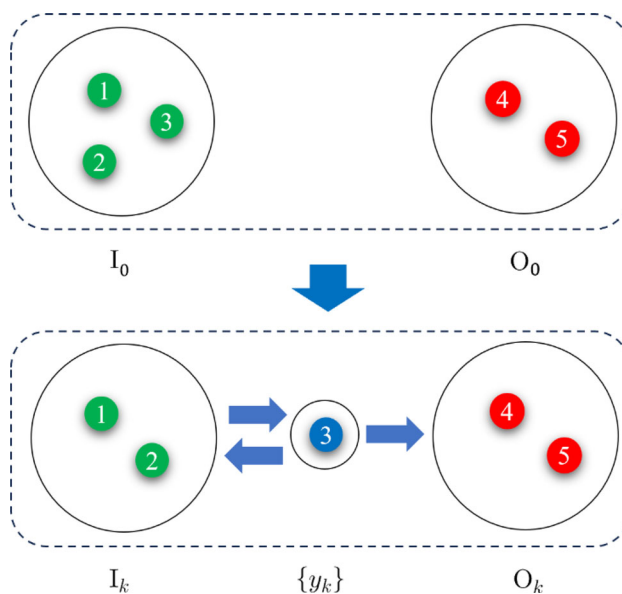


Fig. 4 Procedure of data snooping

critical value k_α . Once the extended w -test failed, which is given as follows:

$$\max_k |w_k| > k_\alpha, y_k \in I_0, \tag{31}$$

then it turns to the identification step.

Identification: the data with the largest test statistic is then identified as an outlier and put into the outlying set.

Adaptation: the LS is implemented for parameter estimation using the data in the inlying set.

3.4 Data refining

Similarly, assuming there is at most one inlier in the initial outlying set O_0 , one can then traverse all data in the outlying set and find out this inlier. This procedure is called *data refining*.

The procedure of data refining is given in Fig. 5. Here, according to Eqs. (24) and (25), one can construct a test statistic w_k for each data in the initial outlying set, $y_k \in O_0$. In this case, for each w_k , we have:

$$O_k \cup \{y_k\} = O_0 \text{ and } I_k = I_0. \tag{32}$$

Then, the data with the smallest test statistic is considered the most suspected inlier in O_0 . Likewise, data refining consists of the following three parts.

Detection: with a significance level α , the significance of the smallest w -test statistic is tested by comparing it with the

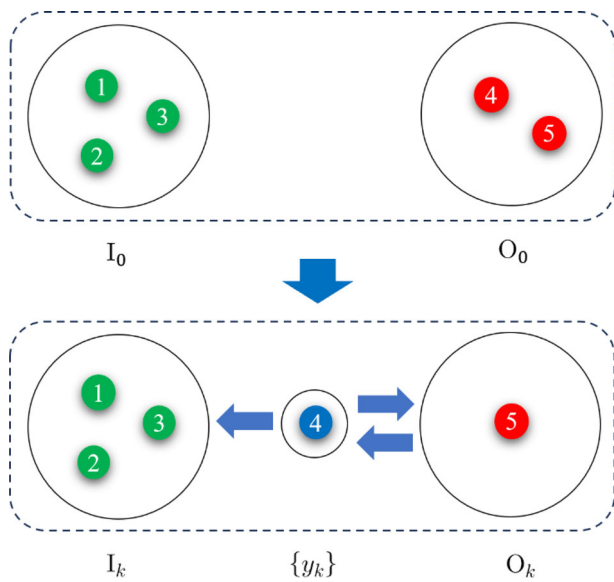


Fig. 5 Procedure of data refining

critical value k_α . Once the w -test is passed, which is given as follows:

$$\min_k |w_k| \leq k_\alpha, y_k \in O_0, \tag{33}$$

then it turns to the identification step.

Identification: the data with the smallest test statistic is then identified as an inlier and put into the inlying set.

Adaptation: the LS is implemented for parameter estimation using the data in the inlying set.

4 Iterative data snooping and iterative data refining

After data division, data snooping and data refining can be employed to find a single outlier or inlier. However, the exact outlier number, denoted as q_* , is often unknown in practical applications. In such cases, given a range of possible outlier numbers based on some knowledge, one can then diagnose outliers within this range. This range is typically defined by setting the minimum and maximum suspected outlier numbers, denoted as q_{\min} and q_{\max} , respectively.

In this section, the iterative forms of data snooping and data refining are presented, which are called IDS and IDR (iterative data refining), respectively. Then, the difference of IDS and IDR are discussed, from the perspective of robustness and accuracy, choice of significance level, and computation cost.

4.1 Iterative data snooping (IDS)

If all data are divided into an outlying set and an inlying set during the initialization, then iterative data snooping (IDS) can be considered as a process of picking the data tested as outliers from the inlying set to the outlying set one by one.

As shown in Fig. 6, the initialization of IDS ($t = 0$) is organized as follows. Given the minimum suspected outlier number q_{\min} , then all of the candidate pairs of the inlying-outlying set can be listed as $I_i^{(0)} \cup O_i^{(0)} = \{y_1, \dots, y_m\}$, $i \in \left\{1, \dots, \binom{m}{q_{\min}}\right\}$. According to Eq. (19), one can find the most possible pair of inlying-outlying sets to construct the initial model in the first iteration ($t = 1$):

$$C_0^{(1)} = \operatorname{argmin}_{C_i^{(0)}} \left\{ \mathbf{y}^T \left[\mathbf{M} - \mathbf{M} \mathbf{C}_i^{(0)} \left(\mathbf{C}_i^{(0)T} \mathbf{M} \mathbf{C}_i^{(0)} \right)^{-1} \mathbf{C}_i^{(0)T} \mathbf{M} \right] \mathbf{y} \right\},$$

$$i \in \left\{ 1, \dots, \binom{m}{q_{\min}} \right\} \tag{34}$$

where $C_i^{(0)} \in \mathbb{R}^{m \times q_{\min}}$ is the design matrix generated by the outlying set $O_i^{(0)}$. Note that q_{\min} is usually set as 0 to avoid dropping any inliers in most cases (Kok 1984; Lehmann and Scheffler 2011; Rofatto et al. 2017; Klein et al. 2022)

As shown in Fig. 7, the iteration procedure in IDS is organized as follows. Assuming in the t th iteration, $I_0^{(t)}$ and $O_0^{(t)}$ are the initial inlying and outlying set, respectively, one could then construct test statistics for each data in the inlying set, which are $w_k^{(t)}, y_k \in I_0^{(t)}$. In this case, for each $w_k^{(t)}$, observations are divided into three disjoint parts, outlying set $O_k^{(t)}$, inlying set $I_k^{(t)}$ and testing set $\{y_k\}$, in which the elements numbers are $q_{\min} + t - 1, m - q_{\min} - t$ and 1, respectively. The relationship among them is given by:

$$I_k^{(t)} \cup \{y_k\} = I_0^{(t)} \text{ and } O_k^{(t)} = O_0^{(t)}, \tag{35}$$

According to Eq. (24), the extended w -test statistic is given as follows (Kok 1984; Teunissen 1990, 2000):

$$w_k^{(t)} = \frac{\mathbf{c}_k^{(t)T} \left[\mathbf{M} - \mathbf{M} \mathbf{C}_k^{(t)} \left(\mathbf{C}_k^{(t)T} \mathbf{M} \mathbf{C}_k^{(t)} \right)^{-1} \mathbf{C}_k^{(t)T} \mathbf{M} \right] \mathbf{y}}{\sigma \sqrt{\mathbf{c}_k^{(t)T} \left[\mathbf{M} - \mathbf{M} \mathbf{C}_k^{(t)} \left(\mathbf{C}_k^{(t)T} \mathbf{M} \mathbf{C}_k^{(t)} \right)^{-1} \mathbf{C}_k^{(t)T} \mathbf{M} \right] \mathbf{c}_k^{(t)}}}, \tag{36}$$

where $\mathbf{c}_k^{(t)} \in \mathbb{R}^m$ and $\mathbf{C}_k^{(t)} \in \mathbb{R}^{m \times (q_{\min} + t - 1)}$ are generated by $\{y_k\}$ and $O_k^{(t)}$, respectively. Particularly, if σ is unknown,

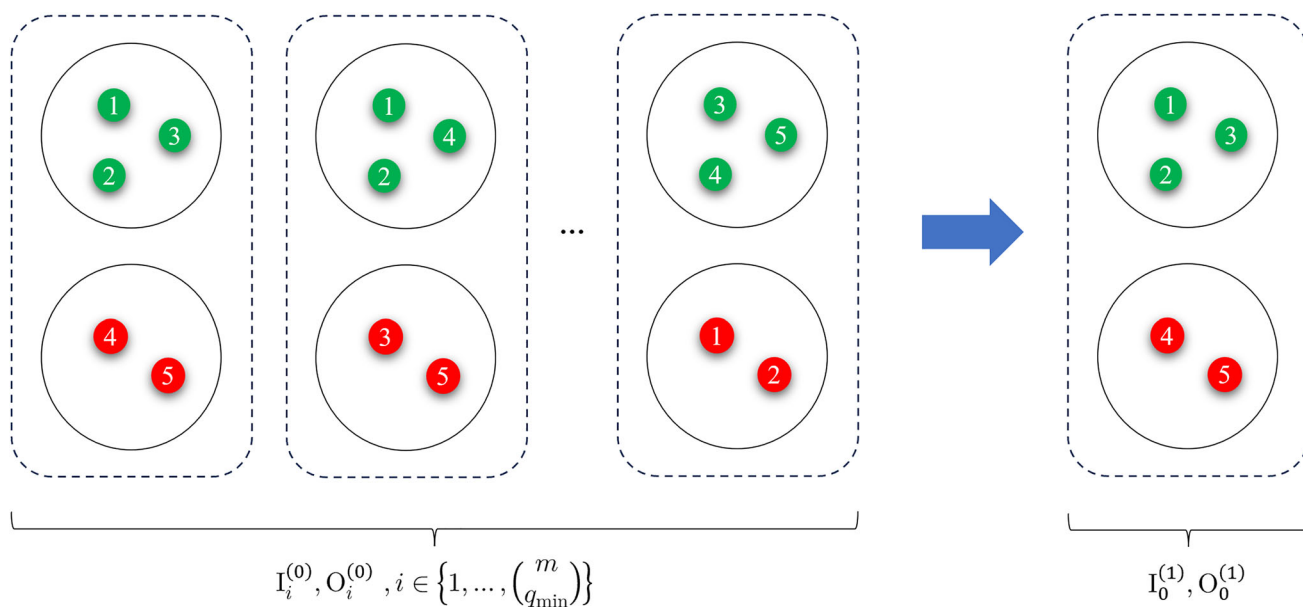


Fig. 6 Initialization procedure of IDS

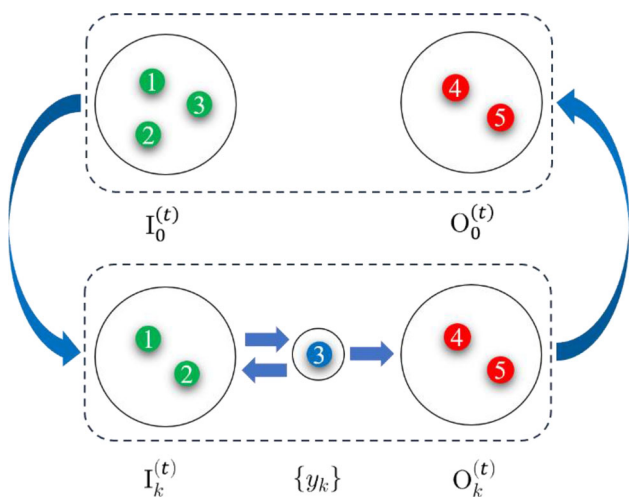


Fig. 7 Iteration procedure in IDS

it can be estimated via Eq. (25) as:

$$\hat{\sigma}_k^{2(t)} = \frac{\mathbf{y}^T \left[\mathbf{M} - \mathbf{M} \mathbf{C}_k^{(t)} \left(\mathbf{C}_k^{(t)T} \mathbf{M} \mathbf{C}_k^{(t)} \right)^{-1} \mathbf{C}_k^{(t)T} \mathbf{M} \right] \mathbf{y}}{m - n - q_{\min} - t}, \tag{37}$$

where $\mathbf{G}_k^{(t)} = \begin{bmatrix} \mathbf{C}_k^{(t)} & \mathbf{c}_k^{(t)} \end{bmatrix}$. The procedure of IDS consists of the following three parts.

Detection: given a significance level α , the largest test statistic is compared to the critical value $k_\alpha^{(t)}$. Once the

extended w -test fails, which is given as follows:

$$\max_k |w_k^{(t)}| > k_\alpha^{(t)}, y_k \in I_0^{(t)}, \tag{38}$$

with

$$k_\alpha^{(t)} = N_{1-\frac{\alpha}{2}}(0, 1) \text{ or } t_{1-\frac{\alpha}{2}}(m - n - q_{\min} - t, 0), \tag{39}$$

then it turns to the identification step.

Identification: the data with the largest test statistic in the inlying set is identified as an outlier and put into the outlying set.

Adaptation: when the iteration is terminated, the LS is implemented for parameter estimation using the data in the inlying set. The terminating condition is that the data number in the outlying set is equal to the maximum suspected outlier number q_{\max} , or all data in the inlying set are considered inliers. Note that q_{\max} is usually less than $m - n$ to make the parameters estimable.

4.2 Iterative data refining (IDR)

Conversely, if all data are divided into an inlying set and an outlying set during the initialization, then iterative data refining (IDR) is a process of picking the data tested as inliers from the outlying set to the inlying set, one after another.

Likewise, the initialization of IDR ($t = 0$) is shown in Fig. 8. Given the maximum suspected outlier number q_{\max} , then all of the potential pairs of inlying-outlying sets can be listed as $I_i^{(0)} \cup O_i^{(0)} = \{y_1, \dots, y_m\}$, $i \in$

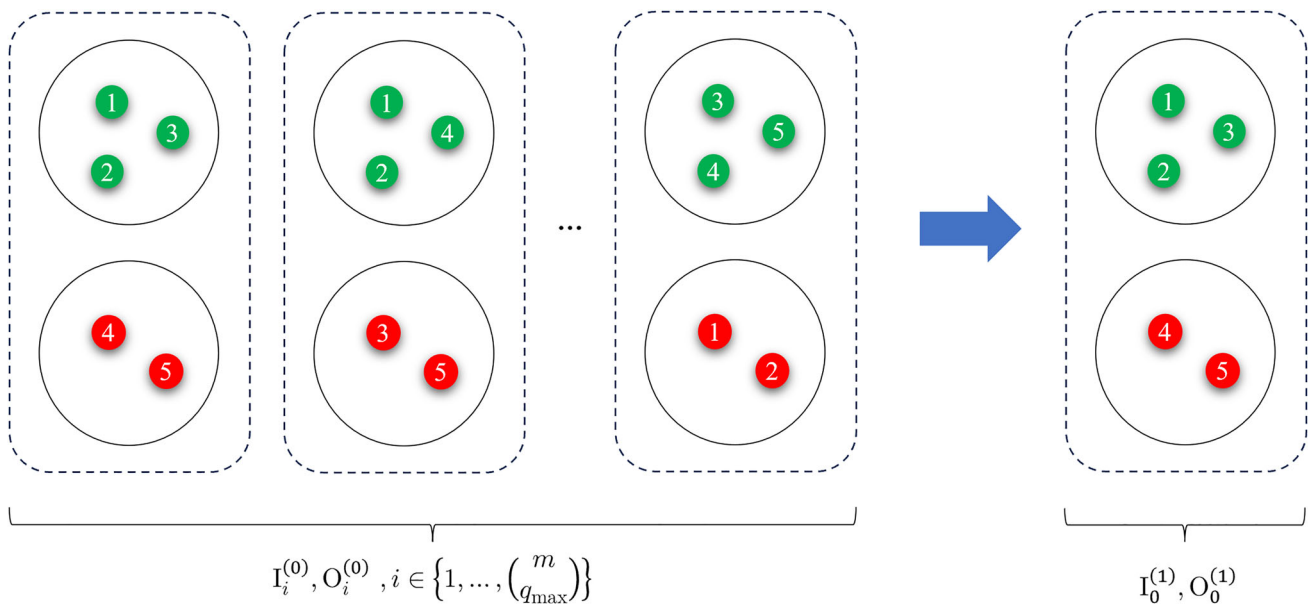


Fig. 8 Initialization procedure of IDR

$\{1, \dots, \binom{m}{q_{\max}}\}$. According to Eq. (19), one can find the most possible pair of inlying-outlying sets to obtain the initial model in the first iteration ($t = 1$):

$$C_0^{(1)} = \underset{C_i^{(0)}}{\operatorname{argmin}} \left\{ \mathbf{y}^T \left[\mathbf{M} - \mathbf{M} \mathbf{C}_i^{(0)} \left(\mathbf{C}_i^{(0)T} \mathbf{M} \mathbf{C}_i^{(0)} \right)^{-1} \mathbf{C}_i^{(0)T} \mathbf{M} \right] \mathbf{y} \right\},$$

$$i \in \left\{ 1, \dots, \binom{m}{q_{\max}} \right\} \quad (40)$$

where $C_i^{(0)} \in \mathbb{R}^m \times q_{\max}$ is the design matrix generated by the outlying set $O_i^{(0)}$. Note that q_{\max} should be less than $m - n$ to make the data division effective.

The iteration procedure in IDR is shown in Fig. 9. Assuming in the t th iteration, $I_0^{(t)}$ and $O_0^{(t)}$ are the initial inlying and outlying sets, respectively, one could then construct test statistics for each data in the outlying set, which are $w_k^{(t)}$, $y_k \in O_0^{(t)}$. For each $w_k^{(t)}$, all data are divided into three disjoint parts, outlying set $O_k^{(t)}$, inlying set $I_k^{(t)}$ and testing set $\{y_k\}$, whose element numbers are $q_{\max} - t$, $m - q_{\max} + t - 1$, and 1, respectively. The relationship among them is given by:

$$O_k^{(t)} \cup \{y_k\} = O_0^{(t)} \text{ and } I_k^{(t)} = I_0^{(t)}, \quad (41)$$

According to Eq. (24), the extended w -test statistic is given by:

$$w_k^{(t)} = \frac{\mathbf{c}_k^{(t)T} \left[\mathbf{M} - \mathbf{M} \mathbf{C}_k^{(t)} \left(\mathbf{C}_k^{(t)T} \mathbf{M} \mathbf{C}_k^{(t)} \right)^{-1} \mathbf{C}_k^{(t)T} \mathbf{M} \right] \mathbf{y}}{\sigma \sqrt{\mathbf{c}_k^{(t)T} \left[\mathbf{M} - \mathbf{M} \mathbf{C}_k^{(t)} \left(\mathbf{C}_k^{(t)T} \mathbf{M} \mathbf{C}_k^{(t)} \right)^{-1} \mathbf{C}_k^{(t)T} \mathbf{M} \right] \mathbf{c}_k^{(t)}}}, \quad (42)$$

where $\mathbf{c}_k^{(t)} \in \mathbb{R}^m$ and $\mathbf{C}_k^{(t)} \in \mathbb{R}^m \times (q_{\max} - t)$ are generated by $\{y_k\}$ and $O_k^{(t)}$, respectively. Particularly, if σ is unknown, it can be estimated via Eq. (25) as:

$$\hat{\sigma}_k^{2(t)} = \frac{\mathbf{y}^T \left[\mathbf{M} - \mathbf{M} \mathbf{G}_k^{(t)} \left(\mathbf{G}_k^{(t)T} \mathbf{M} \mathbf{G}_k^{(t)} \right)^{-1} \mathbf{G}_k^{(t)T} \mathbf{M} \right] \mathbf{y}}{m - n - q_{\max} + t - 1}, \quad (43)$$

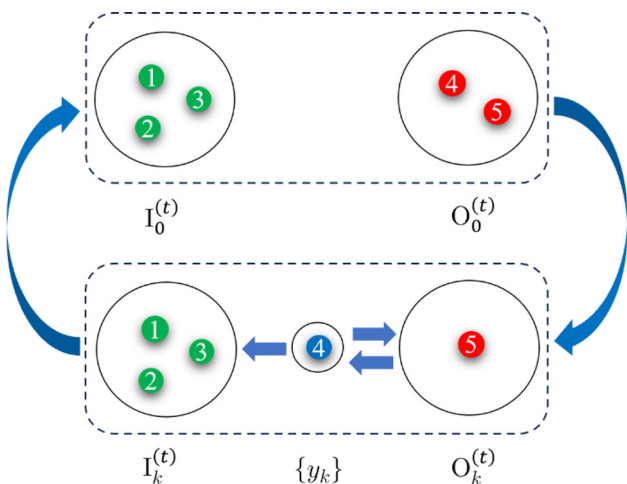


Fig. 9 Iteration procedure in IDR

where $G_k^{(t)} = [C_k^{(t)} \ e_k^{(t)}]$. Similarly, the procedure of IDR consists of the following three parts.

Detection: given a significance level α , the largest test statistic is compared to the critical value $k_\alpha^{(t)}$. Once the extended w -test passes, which is given as follows:

$$\min_k |w_k^{(t)}| \leq k_\alpha^{(t)}, y_k \in O_0^{(t)}, \tag{44}$$

With

$$k_\alpha^{(t)} = N_{1-\frac{\alpha}{2}}(0, 1) \text{ or } t_{1-\frac{\alpha}{2}}(m - n - q_{\max} + t - 1, 0), \tag{45}$$

then it turns to the identification step.

Identification: the data with the smallest test statistic in the outlying set is identified as an inlier and put into the inlying set.

Adaptation: when the iteration is terminated, the LS is implemented for parameter estimation using the data in the inlying set. The terminating condition is that the data number in the outlying set equals the minimum suspected outlier number q_{\min} , or all data in the outlying set are tested as outliers. Likewise, q_{\min} is usually set as 0 to avoid dropping any inliers.

4.3 Precision and robustness

The procedure diagram of IDS and IDR is illustrated in Fig. 10. Generally, for IDS, all data are usually considered inliers and put into the inlying set in the initial stage. In the

following iteration, the extended w -test is used to transfer the data tested as an outlier from the inlying set to the outlying set. This iterative process continues until either the amount of data in the outlying set reaches an upper threshold or there are no more suspected outliers left in the inlying set. Conversely, for IDR, the data are divided into an inlying set and a non-empty outlying set in the initial stage. In the following iteration, the extended w -test is a process of picking the data tested as an inlier from the outlying set to the inlying set. Finally, the process will be terminated either the number of data points in the outlying set reaches a lower threshold, or all data in the outlying set are tested as outliers.

Generally, both IDS and IDR are based on the extended w -test that uses the inlying set to test the suspected data as an inlier or outlier. Therefore, as analyzed in Sect. 3.2, the test performance is up to the number and quality of data in the inlying set. For IDS, the amount of data in the inlying set is sufficient though, the inlying set is usually contaminated in the initial stage in the case of multiple outliers. As a consequence, a contaminated inlying set might invalidate the subsequent test decision thereby compromising the performance of decision and identification for outliers, which is exactly the cause of the masking and swamping effect. Conversely, for IDR, the suspected outliers are moved out of the inlying set as much as possible in the initial stage. This process ensures a more reliable inlying set, thereby mitigating the masking and swamping effect and enhancing the credibility of subsequent tests. Note that, this advantage is based on the premise that the inliers in the observation system are

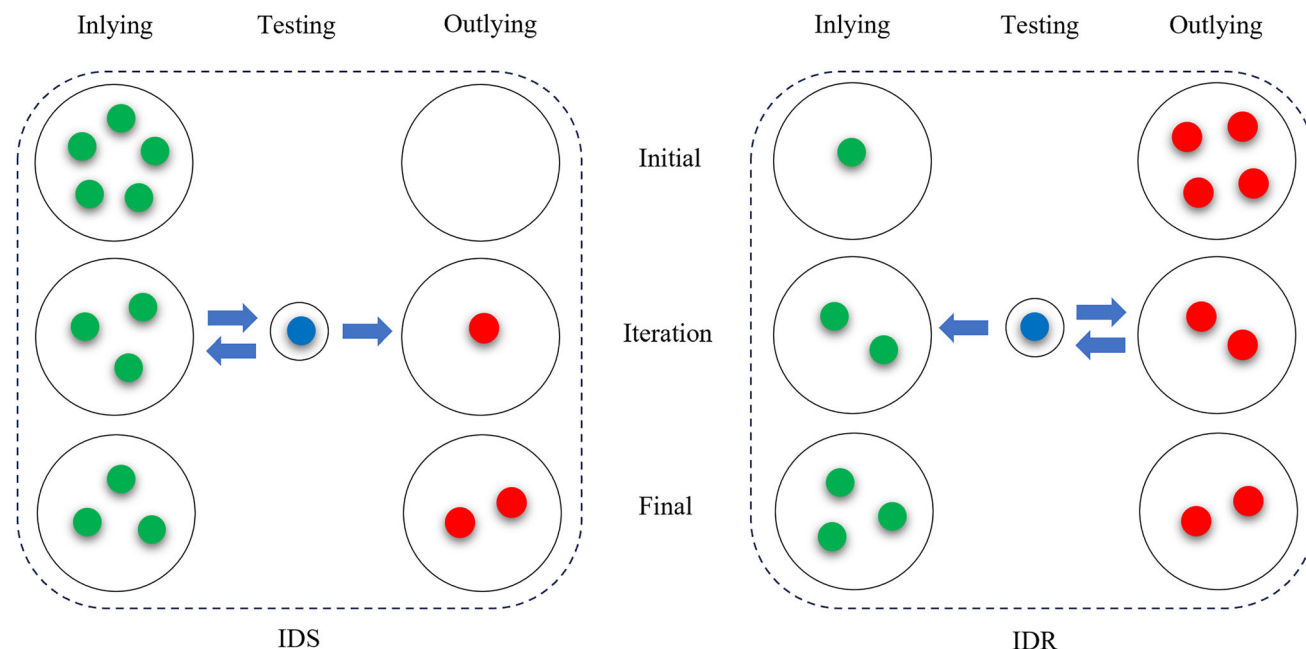


Fig. 10 Procedures of IDS and IDR

sufficient. If the data within the inlying set is severely lacking, the test results may also become untrustworthy due to the low reliability. In conclusion, compared to IDS, IDR will show stronger robustness when dealing with multiple outliers, and might pose a greater risk of precision loss if there is an insufficient amount of data in the inlying set.

4.4 Choice of significance level

Generally, the choice of significance level α for both IDS and IDR should be guided by the principle of controlling the overall false alarm rate, also known as the overall type I error rate. In practice, given an adjustment network, one can first establish the mapping relations between α and the overall false alarm rates. Subsequently, based on these mapping relations, the required α can be determined by specifying an overall false alarm rate. The Monte Carlo Simulation (MCS) proves useful in this process, where each false alarm rate is counted while systematically adjusting α within a specified range. The significance level α corresponding to a given false alarm rate can then be derived (Lehmann 2012; Rofatto et al. 2020b). In particular, if the observation precision σ is known, an overall test can be implemented at the beginning of IDS and IDR to regulate the overall false alarm rate (Kok 1984; Teunissen 2000). Consequently, the subsequent determination of α would become more flexible in this regard. For example, one can choose different α in different iteration through considering the numbers and correlations of the test statistics.

4.5 Computational cost

Both IDS and IDR are computationally expensive when the amount of data is large since there would be quantities of test statistics to be computed. From Eqs. (19) and (24), one can see that the main time-consuming task is to compute the inversion of $C_k^T M C_k$, a symmetric matrix with dimension $q_k \times q_k$. The computational complexity of such an operation is of the order $O(q_k^3)$ (Lehmann and Lösler 2016). Generally, for IDS, there would be $\binom{m}{q_{\min}}$ matrices of dimension $q_{\min} \times q_{\min}$ in the initial stage, and $m - q$ matrices of dimension $q \times q$ during the iteration where q ranges from q_{\min} to $q_{\max} - 1$. As for IDR, there would be $\binom{m}{q_{\max}}$ matrices of dimension $q_{\max} \times q_{\max}$ in the initial stage, and q matrices of dimension $q \times q$ during the iteration where q ranges from q_{\max} to $q_{\min} + 1$. Comparatively, IDR shows more time consumption than IDS, due to the extra computational cost in the initial stage.

5 Example

In this section, an example is used to evaluate the performance of IDR for dealing with outliers. It is useful to elaborate on the theoretical considerations with a simple practical example. Thus, a numerical example of linear fitting with m equidistant data points is given. With error-free abscissae $i = 1, \dots, m$, the observations in Eq. (1) are $y_i = x_1 + ix_2 + e_i$, $e_i \sim N(0, \sigma^2)$, where $m = 10$, $n = 2$ (Lehmann and Lösler 2016). The unknown parameters $x_1 = 0$ and $x_2 = 1$ denote the intercept and slope parameter, respectively. Correspondingly, the design matrix is given by:

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{bmatrix}^T. \quad (46)$$

In addition, the variance–covariance cofactor matrix is given as $Q = I_m$, and the standard deviation of unit weight σ is set as 1.

The numerical example is conducted via MCS. In each simulation, in addition to the normally distributed random errors, different numbers and sizes of gross errors are added to the observations, where the outlier number q_* ranges from 1 to 3, and the gross error size ∇ ranges from 1 to 30 times the observation precision. For fairness, the preset minimum outlier number q_{\min} for IDS and IDR are both set as 0 and the maximum outlier number q_{\max} are both set as 5. Finally, both cases of σ known and unknown are considered for different application scenarios.

In the following discussion, the DIA probability levels (Teunissen 2018; Zaminpardaz and Teunissen 2019; Yang et al. 2021) are used to evaluate the performance of IDS and IDR, which can be calculated via MCS (Hekimoglu and Koch 1999; Rofatto et al. 2020b). Specifically, when there is no outlier, P_{FA} denotes the probability of false alarm (FA) for rejecting any inlier. And when there are some outliers, P_{CD} denotes the probability of correct detection (CD) for rejecting any outlier and P_{CI} denotes the probability of correct identification (CI) for rejecting all outliers. Besides, the fitting root-mean-square error (RMSE) is also used to evaluate the robustness of these methods. The RMSE is calculated by $RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (\hat{y}_i - \bar{y}_i)^2}$, where \hat{y}_i and \bar{y}_i are the estimated value and true value of the observation y_i .

5.1 Control of FA probability

Before the outlier diagnosis, it is very useful to determine the significance level for the statistical test by controlling the P_{FA} or called the type I error rate (Lehmann 2012; Rofatto et al. 2020b). First, Fig. 11 gives the relationship between

P_{FA} and RMSE of IDS, IDR, and LS under different significance levels. Generally, one can see that larger significance levels would lead to higher P_{FA} and RMSE for both IDS and IDR. Specifically, when σ is known, the P_{CA} of IDR are similar to IDS regardless of significance levels. Therefore, the accuracies of IDS and IDR remain at the same level, which is slightly lower than LS. However, in the case of σ unknown, P_{FA} of IDR is larger than that of IDS by using the same significance level. As a consequence, one can see that IDR shows a higher RMSE than LS and IDS. It indicates that IDR might pose a larger risk of precision loss than IDS in case of data insufficiency, especially when a large significance level is chosen. To control the P_{FA} and RMSE, we give the significance levels of IDS and IDR under some

fixed P_{FA} in Table 1. Generally, when σ is known, a fixed P_{FA} corresponds to the same significance level for IDS and IDR. When σ is unknown, the significance level of IDR will be much smaller than that of IDS under the same P_{FA} . In practice, if the σ is known, the overall test can also be used at the beginning of both IDS and IDR to regulate the overall false alarm rate (Kok 1984; Teunissen 2000).

5.2 Comparison of CD, CI probabilities, and RMSE

After the determination of the significance levels in Table 1, the performance of IDS, IDR, and LS are compared in case the observations are contaminated by outliers. First, to evaluate the outlier detection capability of IDS and IDR, P_{CD} of

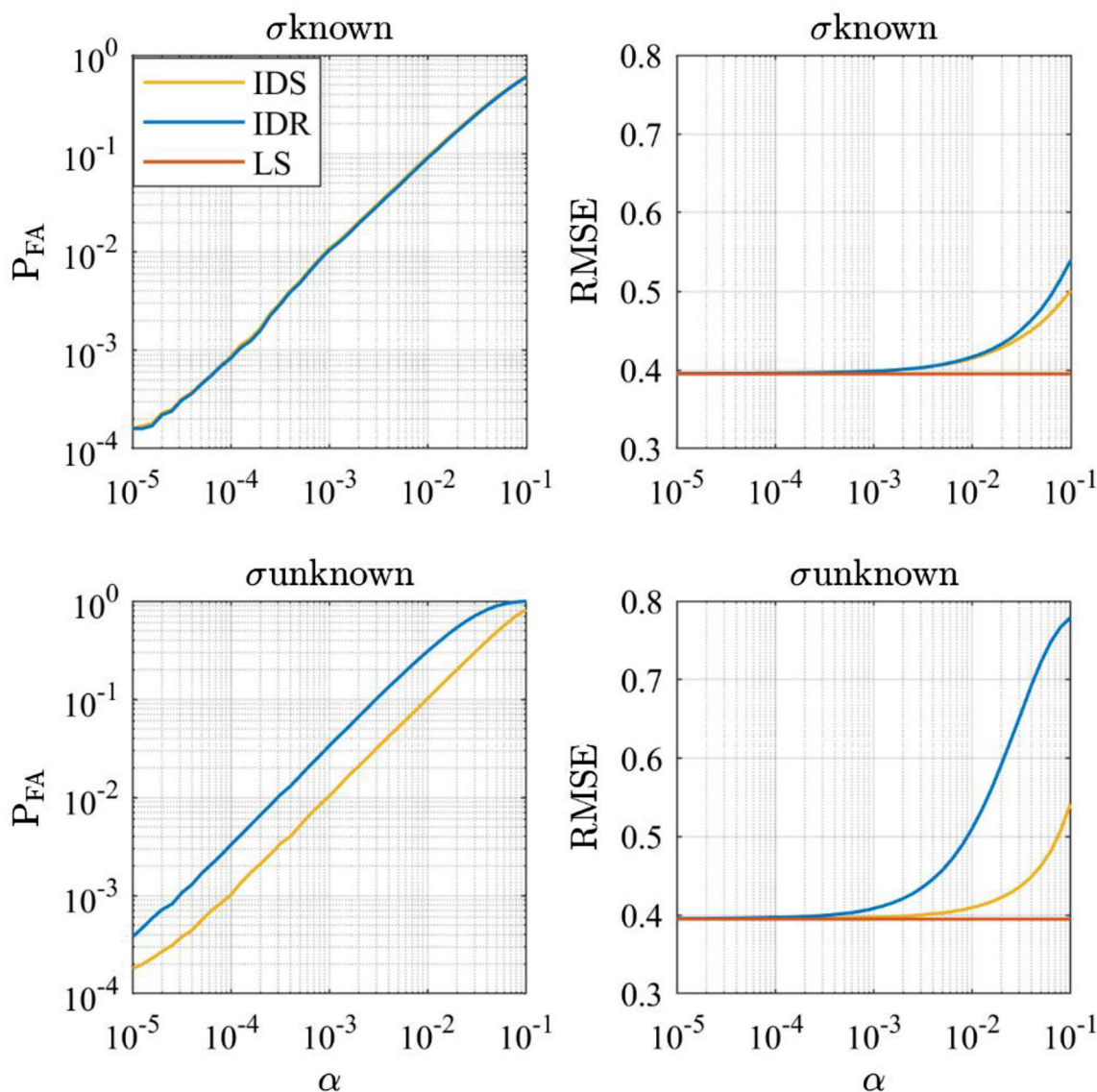


Fig. 11 P_{FA} and RMSE of IDS, IDR, and LS under different significance levels

Table 1 Significance levels of IDS and IDR under different P_{FA}

Methods	$P_{FA} = 1.0 \times 10^{-3}$	$P_{FA} = 1.0 \times 10^{-2}$	$P_{FA} = 1.0 \times 10^{-1}$
IDS (σ known)	1.0×10^{-4}	1.0×10^{-3}	1.0×10^{-2}
IDR (σ known)	1.0×10^{-4}	1.0×10^{-3}	1.0×10^{-2}
IDS (σ unknown)	1.0×10^{-4}	1.0×10^{-3}	1.0×10^{-2}
IDR (σ unknown)	3.2×10^{-5}	3.2×10^{-4}	3.2×10^{-3}

the two methods under different outlier numbers are shown in Fig. 12. In general, one can see that for both IDS and IDR, a larger significance level would lead to a higher P_{CD} , which indicates that although a larger critical value tends to drop more inliers, it also improves the detection capability. In addition, when σ is known, as the gross error size grows,

the P_{CD} of both IDS and IDR show increasing trends and finally keep stable at 100%. However, when σ is unknown, the situation become different. Although IDS and IDR still show comparable P_{CD} in the case of a single outlier, when there are multiple outliers, P_{CD} of IDS shows a downward trend as the gross error size grows even with a great significance level. It is because when there are multiple outliers, the inlying set of IDS is contaminated, and the estimate of σ using Eq. (34) will be biased especially when gross errors are large. As a consequence, this biased estimate will shrink the size of the extended w -test statistics, thereby compromising the detection capability. It proves when σ is unknown the IDS would be severely affected by the masking effect of multiple outliers. Conversely, due to the alleviation of the masking effect, IDR still shows satisfactory performance for outlier detection.

To compare the outlier identification capability of IDS and IDR, Fig. 13 gives P_{CI} of the two methods under different outlier numbers. In general, when the gross error is small, larger

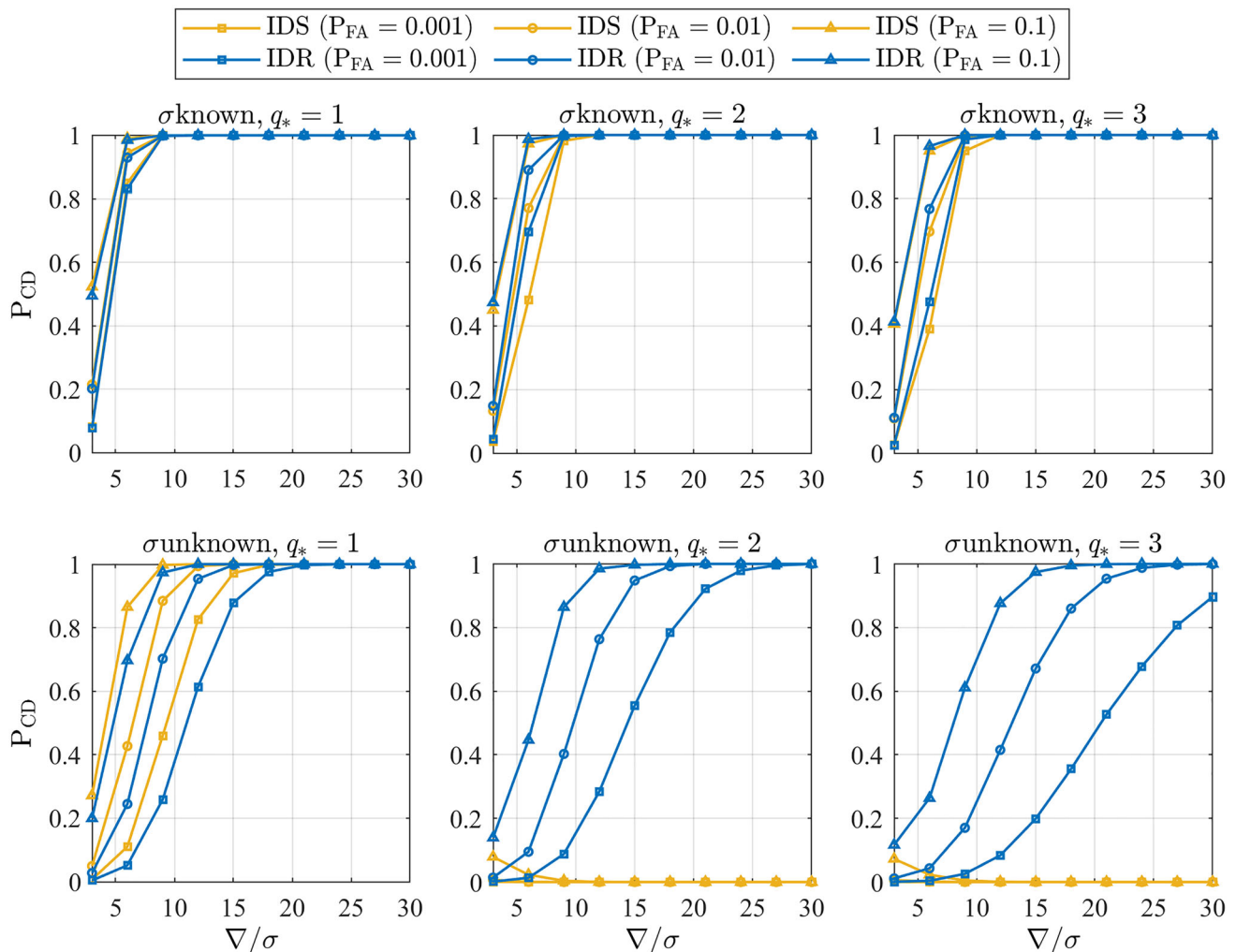


Fig. 12 P_{CD} of IDS and IDR under different significance levels and outlier numbers

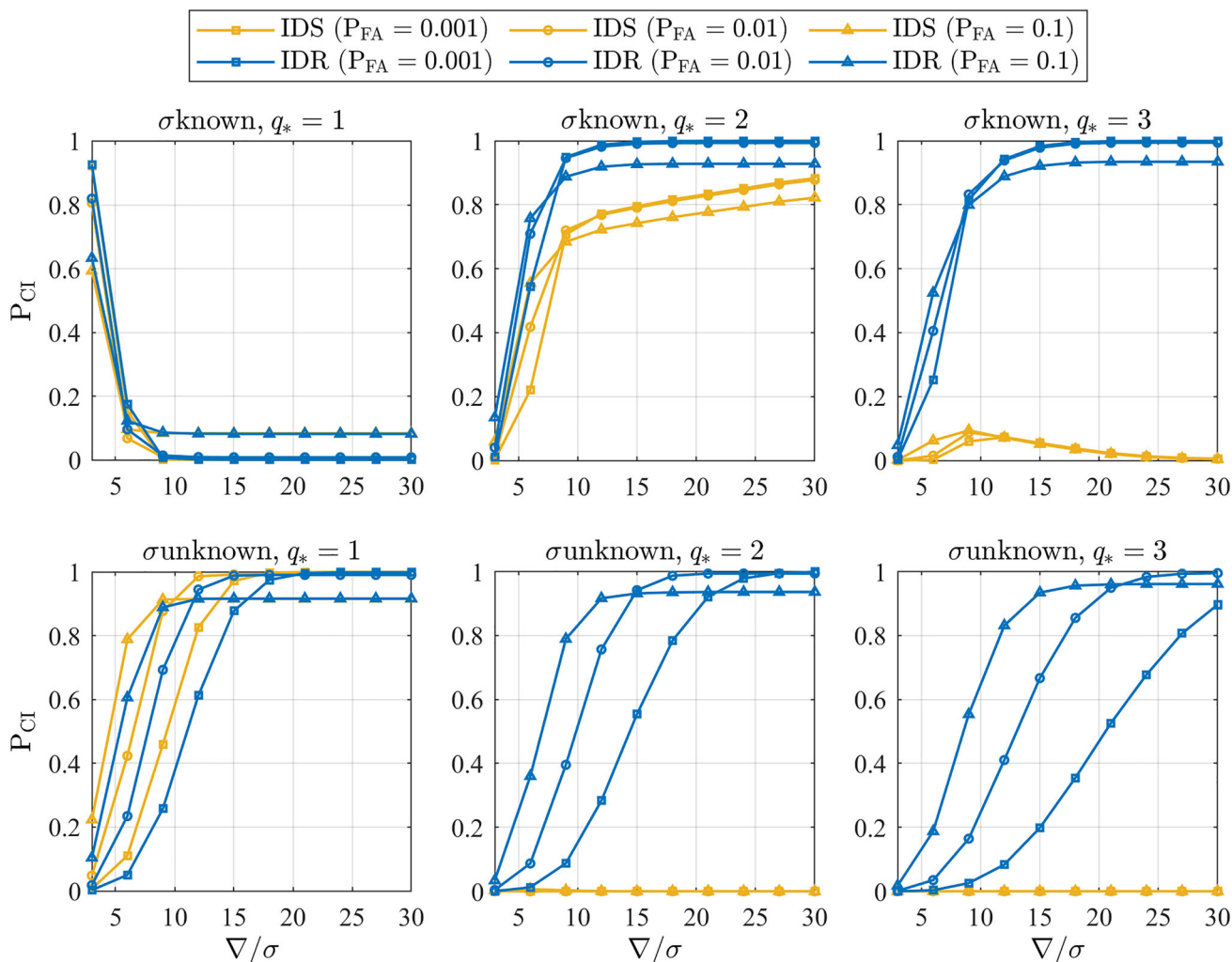


Fig. 13 P_{CI} of IDS and IDR with different significance levels under different outlier numbers

significance levels would bring higher P_{CI} . But it becomes the opposite when the gross error is large. Because in the case of small gross errors, the outlier is hard to detect, and a larger significance level is helpful to detect more outliers. Conversely, in the case of large gross errors, the outlier is easy to detect, and a small significance level will keep more inliers. Specifically, in the case of a single outlier, there is no significant difference between IDS and IDR when σ is known. And when σ is unknown, IDS shows a higher P_{CI} than IDR regardless of the significance level. However, in the case of multiple outliers, no matter σ is known or unknown, IDR shows a notable superiority over IDS on the P_{CI} . Especially, when the outlier number reaches 3, while the P_{CI} of IDR still sees an upward trend with the increase in gross error size, those of IDS almost keep stable at a very low level, which means it is almost impossible for IDS to make a completely correct decision, no matter which significance level is chosen. It indicates that, in case of multiple outliers,

IDR would show stronger potential of identification than IDS due to the alleviation of the masking and swamping effect.

Finally, Fig. 14 compares the fitting RMSEs using IDS, IDR, and LS under different outlier numbers. Generally, for both IDS and IDR, when the gross error size is small, a lower significance level leads to higher accuracy. But when the gross error size is huge, a higher significance level keeps higher robustness. Specifically, when there is only a single outlier, the RMSE of LS dramatically increases as the gross error size enlarges, which means LS is even not robust for rare outliers. In comparison, by using the two outlier diagnosis methods, fitting errors can be generally reduced to the same extent. At this time, IDS keeps slightly higher accuracy than IDR, especially in the case that σ is unknown. However, when there is more than one outlier, although IDS have a certain resistance to multiple outliers, the robustness would severely degrade as the gross error magnitude increases. For example, when σ is known and there are three outliers, IDS shows even lower accuracy than LS. Because although the outliers can

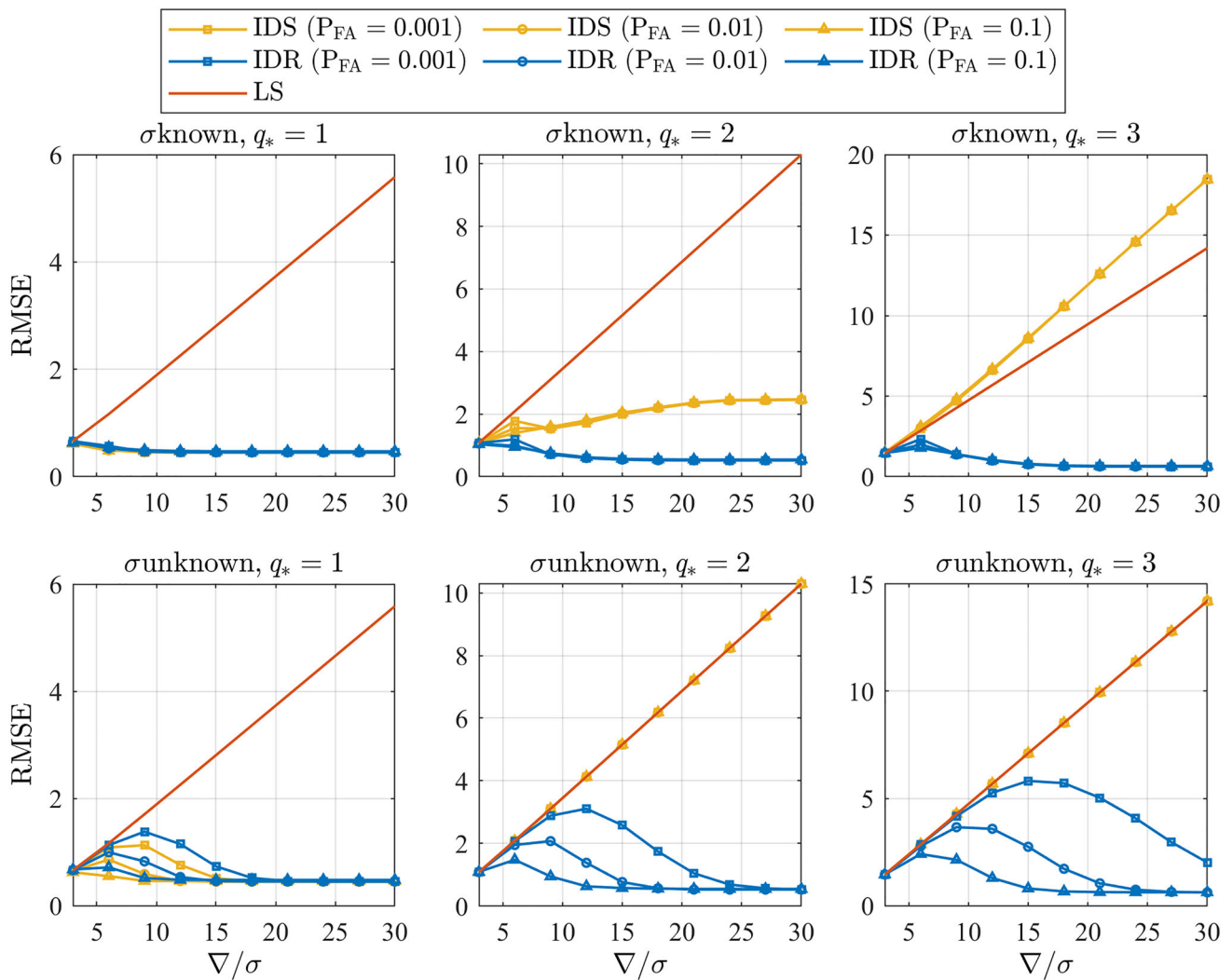


Fig. 14 RMSE of LS, IDS, and IDR with different significance levels under different outlier numbers

be detected by IDS in this case (see Fig. 12), many inliers will also be rejected due to the low identify capability (see Fig. 13). In addition, when σ is unknown, IDS could hardly detect any outlier (see Fig. 12), thereby keeping the same performance as LS. This verifies that the IDS, although has been widely used in various practical applications, actually may lose robustness for handling multiple outliers with remarkable sizes. In comparison, the fitting errors of using the IDR are generally much smaller and can always be controlled, even in the case of gross errors with large numbers and magnitude. It reveals that due to the alleviation of the masking and swamping effect, IDR would show higher robustness than IDS when dealing with multiple outliers.

5.3 Analysis of suspected outlier numbers

In this subsection, the influence of q_{\max} on the performance of IDS and IDR are analyzed. For fairness, the P_{FA} of both

IDS and IDR are chosen as 1.0×10^{-2} and the gross error is set as 15σ . First, Fig. 15 shows the significance level of IDS and IDR under different q_{\max} . Generally, when σ is known, the significance levels of IDS and IDR under a fixed P_{FA} will remain consistent and unchanged regardless of the choice of q_{\max} . In case that σ is unknown, the significance level of IDS is still unchanged as the q_{\max} grows. However, the significance level of IDR is much slower than that of IDS and shows a downward trend as the increase in q_{\max} . It indicates that for IDR, while the significance level will not be influenced by q_{\max} in the case of σ known, with a larger q_{\max} , a smaller significance level is needed to control P_{FA} in the case of σ unknown.

Using these determined significance levels, the RMSE of IDS, IDR, and LS under a fixed P_{FA} and different q_{\max} are shown in Fig. 16. Note that the q_{\max} , which ranges from 3 to 7, is always larger than q_* to guarantee the robustness. Generally, when there is only a single outlier, IDS and IDR

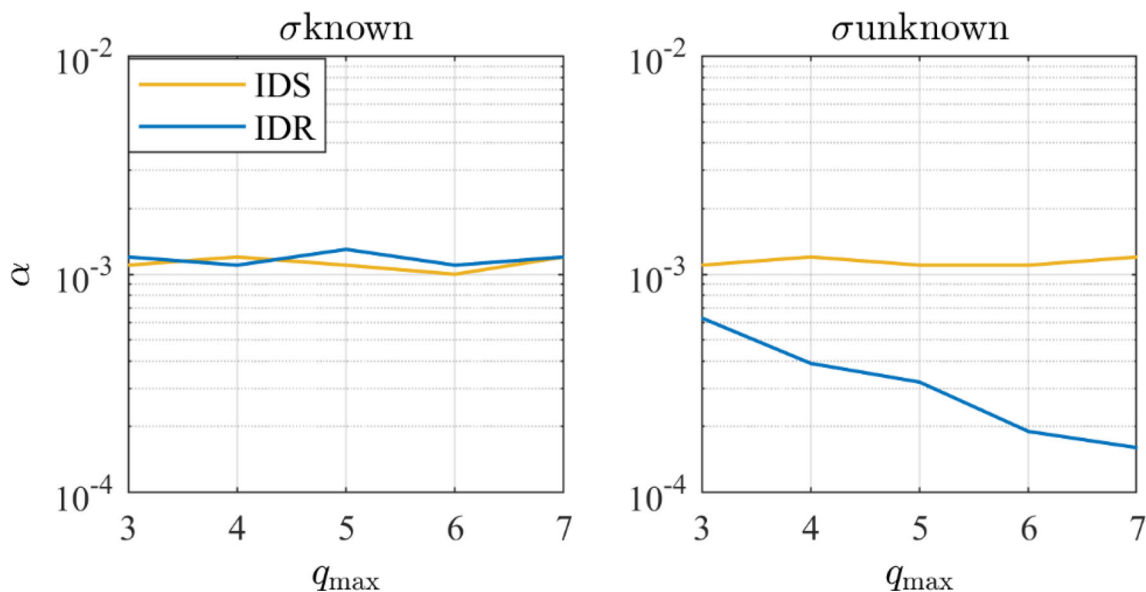


Fig. 15 Significance level of IDS and IDR under a fixed P_{FA} and different q_{max}

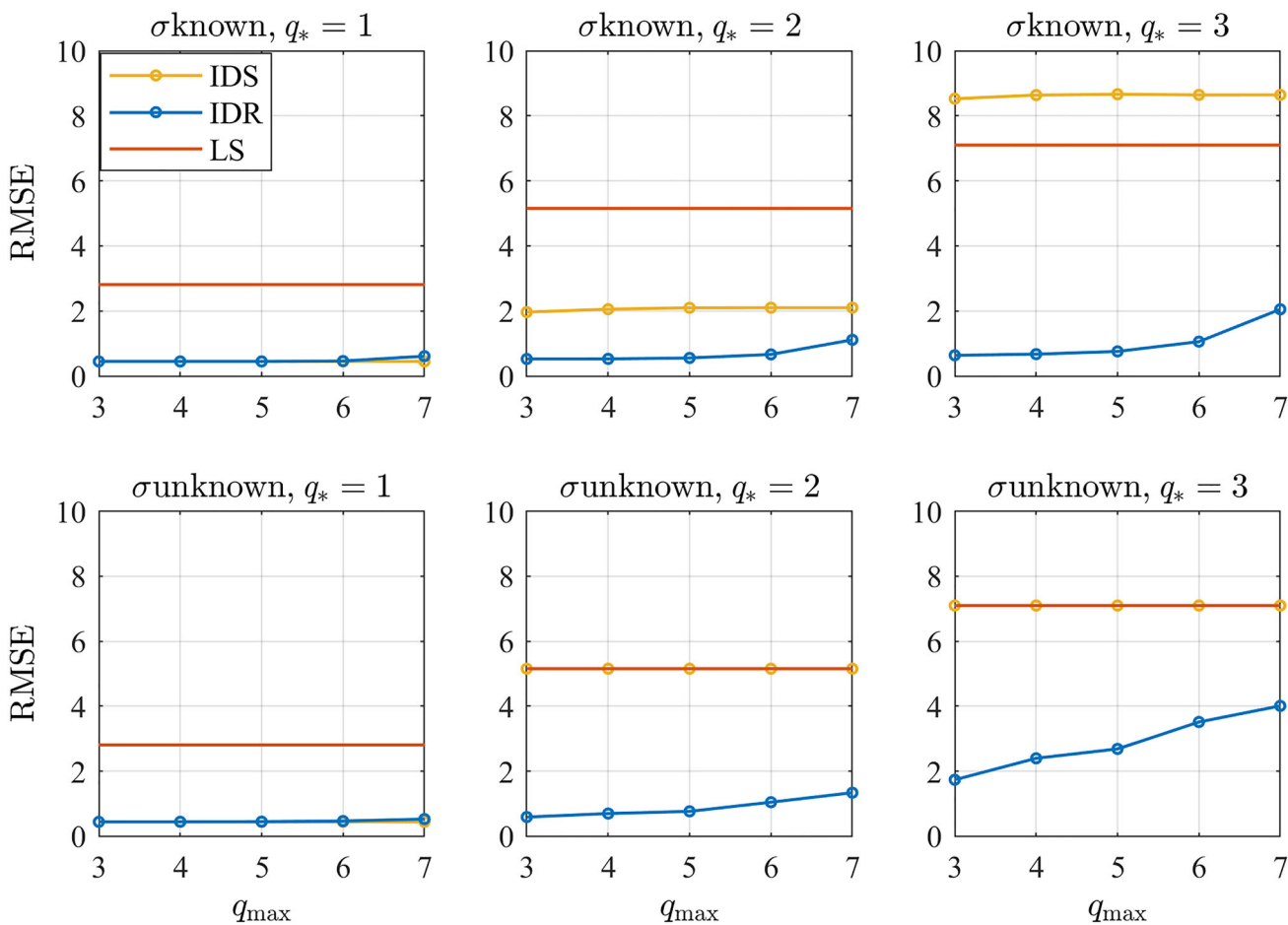


Fig. 16 RMSE of IDS, IDR, and LS under a fixed P_{FA} and different q_{max}

always show higher improvement than LS on robustness. Comparatively, IDS keeps slightly higher accuracy than IDR, especially when a large q_{\max} is chosen. In the case of multiple outliers, while the robustness of IDS severely degrades, IDR always shows higher robustness due to the alleviation of the masking and swamping effect. In addition, compared to IDS, the performance of IDR is more sensitive to the choice of q_{\max} especially when σ is unknown. In other words, the RMSE of IDR always shows an increasing trend as the q_{\max} grows, since that a larger q_{\max} would lead to greater risk of dropping inliers. It reveals that under the premise that q_{\max} is greater than the outlier number, a smaller q_{\max} is more popular to ensure the accuracy of parameter estimates.

6 Conclusion

In this contribution, the causes of masking and swamping effects are investigated, and a new method of outlier diagnostics is proposed to alleviate these phenomena. First, according to the concept of data division, an extended form of the w -test with its associated reliability measure is presented. Secondly, based on the extended w -test, both data snooping and IDS are reinterpreted theoretically. Finally, a new outlier diagnostic method and its iterative form are proposed, which are called data refining and iterative data refining, respectively. While data snooping is a process of snooping outliers from a preset inlying set to the outlying set, data refining can be considered as a reverse process to refine inliers from an outlying set to the inlying one.

A linear fitting example is used to evaluate the performance of the proposed IDR when dealing with outliers. Generally, when there is a single outlier, IDR shows similar performances with IDS on both probabilities of correct decision and accuracy of the parameter estimate. However, when the outlier number grows, the correct decision probability and estimation accuracy of IDS will degrade dramatically due to the masking and swamping effect. Conversely, IDR still maintains a stable and satisfactory performance, even in the case of gross errors with large numbers and magnitudes. It proves that IDR outperforms IDS when dealing with multiple outliers, due to the alleviation of masking and swamping effect.

It should be noted that compared to IDS, the applications of IDR still face several challenges. First, IDR poses a larger risk of precision loss for parameter estimation, especially when the observation precision is unknown. Therefore, it is especially important to control the false alarm rate to avoid dropping too many inliers. Second, the performance of IDR is more sensitive to the preset parameter q_{\max} . In other words, a larger q_{\max} usually causes lower precision of parameter estimation, once it exceeds the truth outlier number. Factors to consider when choosing q_{\max} include data redundancy, the

observation environment, and the task requirements. Finally, IDR usually costs relatively higher computing time because finding a suitable model in the initial stage is usually computationally expensive. Therefore, how to find the most reliable and efficient initialization method will be further investigated in our future works.

Acknowledgements This work is sponsored by the National Natural Science Foundation of China (42274030, 42192532) and the Fundamental Research Funds for the Central Universities (22120210522).

Author contribution Y.Y. proposed the key idea, designed the research, and wrote the paper draft; L.Y. and Y.S. supervised the research and revised the manuscript.

Funding The National Natural Science Foundation of China (42274030, 42192532) and the Fundamental Research Funds for the Central Universities (22120210522).

Data availability Necessary data is accessible in the example section of the manuscript.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Appendix of Proof

Proof of Eqs. (22)-(23)

According to Eq. (21), the normal equation can be given by:

$$\begin{bmatrix} \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A} & \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{G}_k \\ \mathbf{G}_k^T \mathbf{Q}^{-1} \mathbf{A} & \mathbf{G}_k^T \mathbf{Q}^{-1} \mathbf{G}_k \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_k \\ \hat{\mathbf{v}}_k \end{bmatrix} = \begin{bmatrix} \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{y} \\ \mathbf{G}_k^T \mathbf{Q}^{-1} \mathbf{y} \end{bmatrix}. \quad (\text{A1})$$

where

$$\mathbf{G}_k = [\mathbf{C}_k \ \mathbf{c}_k], \quad \nabla_k = [\mathbf{b}_k^T \ \nabla_k^T]^T. \quad (\text{A2})$$

Here, we apply a Gaussian elimination. The above equation is pre-multiplied with the following square and full-rank matrix:

$$\begin{bmatrix} \mathbf{I}_n & 0 \\ -\mathbf{G}_k^T \mathbf{Q}^{-1} \mathbf{A} (\mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A})^{-1} & \mathbf{I}_{q_{k+1}} \end{bmatrix}, \quad (\text{A3})$$

resulting in:

$$\begin{bmatrix} \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A} & \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{G}_k \\ 0 & \mathbf{G}_k^T [\mathbf{Q}^{-1} - \mathbf{Q}^{-1} \mathbf{A} (\mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Q}^{-1}] \mathbf{G}_k \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_k \\ \hat{\mathbf{v}}_k \end{bmatrix} = \begin{bmatrix} \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{y} \\ \mathbf{G}_k^T [\mathbf{Q}^{-1} - \mathbf{Q}^{-1} \mathbf{A} (\mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Q}^{-1}] \mathbf{y} \end{bmatrix}. \quad (\text{A4})$$

Considering that $M = Q^{-1} - Q^{-1}A(A^T Q^{-1}A)^{-1}A^T Q^{-1}$, so we have:

$$G_k^T M G_k \hat{V}_k = G_k^T M y. \tag{A5}$$

Inserting Eq. (A2) into Eq. (A5), we have:

$$\begin{bmatrix} C_k^T M C_k & C_k^T M c_k \\ c_k^T M C_k & c_k^T M c_k \end{bmatrix} \begin{bmatrix} \hat{b}_k \\ \hat{V}_k \end{bmatrix} = \begin{bmatrix} C_k^T M y \\ c_k^T M y \end{bmatrix}. \tag{A6}$$

Likewise, the above equation is pre-multiplied with the following square and full-rank matrix:

$$\begin{bmatrix} I_{q_k} & 0 \\ -c_k^T M C_k (C_k^T M C_k)^{-1} & 1 \end{bmatrix}, \tag{A7}$$

resulting in:

$$\begin{bmatrix} C_k^T M C_k & C_k^T M c_k \\ 0 & c_k^T M c_k - c_k^T M C_k (C_k^T M C_k)^{-1} C_k^T M c_k \end{bmatrix} \begin{bmatrix} \hat{b}_k \\ \hat{V}_k \end{bmatrix} = \begin{bmatrix} C_k^T M y \\ c_k^T M y - c_k^T M C_k (C_k^T M C_k)^{-1} C_k^T M y \end{bmatrix}. \tag{A8}$$

Therefore, we have:

$$\begin{aligned} & c_k^T \left[M - M C_k (C_k^T M C_k)^{-1} C_k^T M \right] c_k \hat{V}_k \\ &= c_k^T \left[M - M C_k (C_k^T M C_k)^{-1} C_k^T M \right] y, \end{aligned} \tag{A9}$$

which indicates that:

$$\hat{V}_k = \frac{c_k^T \left[M - M C_k (C_k^T M C_k)^{-1} C_k^T M \right] y}{c_k^T \left[M - M C_k (C_k^T M C_k)^{-1} C_k^T M \right] c_k}, \tag{A10}$$

and

$$q_{\hat{V}_k}^2 = \frac{1}{c_k^T \left[M - M C_k (C_k^T M C_k)^{-1} C_k^T M \right] c_k}. \tag{A11}$$

The proof is done.

Proof of Eq. (25)

Following Eq. (A5), we have:

$$\hat{V}_k = \left(G_k^T M G_k \right)^{-1} G_k^T M y. \tag{A12}$$

Inserting Eq. (A12) into (A4), we have:

$$A^T Q^{-1} A \hat{x}_k + A^T Q^{-1} G_k \left(G_k^T M G_k \right)^{-1} G_k^T M y = A^T Q^{-1} y. \tag{A13}$$

Solving Eq. (A13), results in:

$$\hat{x}_k = \left(A^T Q^{-1} A \right)^{-1} A^T Q^{-1} \left[I_m - G_k \left(G_k^T M G_k \right)^{-1} G_k^T M \right] y. \tag{A14}$$

Then, inserting Eqs. (A12) and (A14) into (21), we have:

$$\hat{e}_k = \left\{ I_m - A \left(A^T Q^{-1} A \right)^{-1} A^T Q^{-1} \left[I_m - G_k \left(G_k^T M G_k \right)^{-1} G_k^T M \right] \right\} y. \tag{A15}$$

Then, the estimate of σ can be given by:

$$\begin{aligned} \hat{\sigma}_k &= \sqrt{\frac{\hat{e}_k^T Q^{-1} \hat{e}_k}{m - n - q_k - 1}} \\ &= \sqrt{\frac{y^T \left[M - M G_k \left(G_k^T M G_k \right)^{-1} G_k^T M \right] y}{m - n - q_k - 1}}. \end{aligned} \tag{A16}$$

The proof is done.

References

Alberda J (1976) Quality control in surveying. *Chart Surv* 4(2):23–28
 Anscombe FJ (1960) Rejection of outliers. *Technometrics* 2:123–146. <https://doi.org/10.1080/00401706.1960.10489888>
 Baarda W (1967) Statistical concepts in geodesy. Netherlands Geodetic Commission Publication on Geodesy, Delft
 Baarda W (1968) A testing procedure for use in geodetic networks. Netherlands Geodetic Commission Publication on Geodesy, Delft
 Barnett V, Lewis T (1978) Outliers in statistical data. Wiley, New York
 Beckman RJ, Cook RD (1983) Outlier s. *Technometrics* 25:119–149. <https://doi.org/10.1080/00401706.1983.10487840>
 Box GEP (1953) Non-normality and tests on variances. *Biometrika* 40:318–335
 Daniel C (1960) Locating outliers in factorial experiments. *Technometrics* 2:149–156. <https://doi.org/10.1080/00401706.1960.10489889>
 David HA, Paulson AS (1965) The performance of several tests for outliers. *Biometrika* 52:429–436. <https://doi.org/10.2307/2333695>
 Donoho DL, Huber PJ (1983) The notion of breakdown point. A festschrift for Erich L Lehmann. Wadsworth, Belmont, p 157184
 Duchnowski R (2010) Median-based estimates and their application in controlling reference mark stability. *J Surv Eng* 136:47–52. [https://doi.org/10.1061/\(ASCE\)SU.1943-5428.0000014](https://doi.org/10.1061/(ASCE)SU.1943-5428.0000014)
 Duchnowski R (2013) Hodges–Lehmann estimates in deformation analyses. *J Geod* 87:873–884. <https://doi.org/10.1007/s00190-013-0651-2>
 Edgeworth FY (1887) On observations relating to several quantities. *Hermathena* 6:279–285
 Ellenberg JH (1973) The joint distribution of the standardized least squares residuals from a general linear regression. *J Am Stat Assoc* 68:941–943. <https://doi.org/10.1080/01621459.1973.10481450>

- Ellenberg JH (1976) Testing for a single outlier from a general linear regression. *Biometrics* 32:637. <https://doi.org/10.2307/2529752>
- Ferguson TS (1961) On the rejection of outliers. University of California Press, Berkeley, pp 253–287
- Fieller N (1976) Some problems related to the rejection of outlying observations. University of Sheffield, Sheffield
- Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM* 24:381–395
- Fisher RA (1925) Theory of statistical estimation. *Math Proc Camb Philos Soc* 22:700–725. <https://doi.org/10.1017/S0305004100009580>
- Förstner W (1983) Reliability and discemability of extended Gauss-Markov models. Seminar on mathematical models to outliers and systematic errors. Deutsche Geodätische Kommission, Munich, Germany, pp 79–103
- Galpin JS, Hawkins DM (1981) Rejection of a single outlier in two- or three-way layouts. *Technometrics* 23:65–70. <https://doi.org/10.1080/00401706.1981.10486238>
- Gentle JE (1978) Testing for outliers in linear regression. Contributions to survey sampling and applied statistics. Elsevier, Amsterdam, pp 223–233
- Grubbs FE (1969) Procedures for detecting outlying observations in samples. *Technometrics* 11:1–21. <https://doi.org/10.1080/00401706.1969.10490657>
- Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA (eds) (1986) Robust statistics: the approach based on influence functions. Wiley, New York
- Hawkins DM (1980) Identification of outliers. Springer, Netherlands, Dordrecht
- Hekimoğlu Ş (1997) Finite sample breakdown points of outlier detection procedures. *J Surv Eng* 123:15–31. [https://doi.org/10.1061/\(ASCE\)0733-9453\(1997\)123:1\(15\)](https://doi.org/10.1061/(ASCE)0733-9453(1997)123:1(15))
- Hekimoğlu Ş (1999) Robustifying conventional outlier detection procedures. *J Surv Eng* 125:69–86. [https://doi.org/10.1061/\(ASCE\)0733-9453\(1999\)125:2\(69\)](https://doi.org/10.1061/(ASCE)0733-9453(1999)125:2(69))
- Hekimoğlu S, Koch K (1999) How can reliability of the robust methods be measured? Third Turkish-German joint geodetic days: towards a digital age. Istanbul Technical University, Istanbul, Turkey, pp 179–196
- Hodges JL Jr (1967) Efficiency in normal samples and tolerance of extreme values for some estimates of location. University of California Press, Berkeley, pp 163–186
- Hodges JL, Lehmann EL (1963) Estimates of location based on rank tests. In: Rojo J (ed) Selected works of E. L. Lehmann. Springer, Boston, MA, pp 287–300
- Holland PW, Welsch RE (1977) Robust regression using iteratively reweighted least-squares. *Commun Stat Theory Methods* 6:813–827. <https://doi.org/10.1080/03610927708827533>
- Huber PJ (1964) Robust estimation of a location parameter. In: Kotz S, Johnson NL (eds) Breakthroughs in statistics. Springer, New York, NY, pp 492–518
- Jaekel LA (1972) Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann Math Stat* 43:1449–1458
- Khodabandeh A, Amiri-Simkooei AR (2011) Recursive algorithm for L1 norm estimation in linear models. *J Surv Eng* 137:1–8. [https://doi.org/10.1061/\(ASCE\)SU.1943-5428.0000031](https://doi.org/10.1061/(ASCE)SU.1943-5428.0000031)
- Klein I, Suraci SS, de Oliveira LC et al (2022) An attempt to analyse iterative data snooping and L1-norm based on Monte Carlo simulation in the context of leveling networks. *Surv Rev* 54:70–78. <https://doi.org/10.1080/00396265.2021.1878338>
- Koch K-R (1999) Parameter estimation and hypothesis testing in linear models. Springer, Berlin Heidelberg
- Koch KR (2013) Robust estimation by expectation maximization algorithm. *J Geod* 87:107–116. <https://doi.org/10.1007/s00190-012-0582-3>
- Koch K-R (2015) Minimal detectable outliers as measures of reliability. *J Geod* 89:483–490. <https://doi.org/10.1007/s00190-015-0793-5>
- Koenker R, Hallock KF (2001) Quantile regression. *J Econ Perspect* 15:143–156. <https://doi.org/10.1257/jep.15.4.143>
- Kok J (1984) On data snooping and multiple outlier testing. NOAA technical report NOS NGS 30. US Department of Commerce National Oceanic and Atmospheric Administration National Ocean Service Charting and Geodetic Service, Washington, DC
- Krarpup T, Kubik K, Juhl J (1980) Gotterdammerung over least squares adjustment. In: Proceedings of international society for photogrammetry 14th congress, vol 3, pp 370–378
- Lehmann R (2012) Improved critical values for extreme normalized and studentized residuals in Gauss–Markov models. *J Geod* 86:1137–1146. <https://doi.org/10.1007/s00190-012-0569-0>
- Lehmann R (2013) On the formulation of the alternative hypothesis for geodetic outlier detection. *J Geod* 87:373–386. <https://doi.org/10.1007/s00190-012-0607-y>
- Lehmann R, Lösler M (2016) Multiple outlier detection: hypothesis tests versus model selection by information criteria. *J Surv Eng* 142:04016017. [https://doi.org/10.1061/\(ASCE\)SU.1943-5428.0000189](https://doi.org/10.1061/(ASCE)SU.1943-5428.0000189)
- Lehmann R, Scheffler T (2011) Monte Carlo-based data snooping with application to a geodetic network. *J Appl Geod*. <https://doi.org/10.1515/JAG.2011.014>
- McMillan RG (1971) Tests for one or two outliers in normal samples with unknown variance. *Technometrics* 13:87–100. <https://doi.org/10.1080/00401706.1971.10488756>
- Mickey MR, Jean Dunn O, Clark V (1967) Note on the use of stepwise regression in detecting outliers. *Comput Biomed Res* 1:105–111. [https://doi.org/10.1016/0010-4809\(67\)90009-2](https://doi.org/10.1016/0010-4809(67)90009-2)
- Nair KR (1948) The distribution of the extreme deviate from the sample mean and its studentized form. *Biometrika* 35:118–144. <https://doi.org/10.2307/2332634>
- Pearson ES, Sekar CC (1936) The efficiency of statistical tools and a criterion for the rejection of outlying observations. *Biometrika* 28:308–320. <https://doi.org/10.2307/2333954>
- Pope AJ (1976) The statistics of residuals and the detection of outliers. NOAA technical report NOS 65 NGS 1. U.S. National Geodetic Survey, Washington, DC
- Quesenberry CP, David HA (1961) Some tests for outliers. *Biometrika* 48:379–390. <https://doi.org/10.2307/2332759>
- Rofatto VF, Matsuoka MT, Klein I (2017) An attempt to analyse baarda's iterative data snooping procedure based on Monte Carlo simulation. *SA J Geom* 6:416. <https://doi.org/10.4314/sajg.v6i3.11>
- Rofatto VF, Matsuoka MT, Klein I et al (2020a) A half-century of Baarda's concept of reliability: a review, new perspectives, and applications. *Surv Rev* 52:261–277. <https://doi.org/10.1080/00396265.2018.1548118>
- Rofatto VF, Matsuoka MT, Klein I et al (2020b) A monte carlo-based outlier diagnosis method for sensitivity analysis. *Remote Sens* 12:860. <https://doi.org/10.3390/rs12050860>
- Rosner B (1975) On the detection of many outliers. *Technometrics* 17:221. <https://doi.org/10.2307/1268354>
- Rousseeuw PJ (1984) Least median of squares regression. *J Am Stat Assoc* 79:871–880. <https://doi.org/10.1080/01621459.1984.10477105>
- Rousseeuw PJ, Leroy AM (1987) Robust regression and outlier detection. Wiley, New York
- Rousseeuw PJ, Yohai V (1984) Robust regression by means of S-estimators. In: Franke J, Härdle W, Martin D (eds) Robust and nonlinear time series analysis. Springer, New York, NY, USA, pp 256–272
- Sarhan AE, Greenberg BG (1956) Estimation of location and scale parameters by order statistics from singly and doubly censored

- samples. *Ann Math Stat* 27:427–451. <https://doi.org/10.1214/aoms/1177728267>
- Srikantan KS (1961) Testing for the single outlier in a regression model. *Sankhyā Indian J Stat Ser A* 23:251–260
- Stefansky W (1972) Rejecting outliers in factorial designs. *Technometrics* 14:469–479. <https://doi.org/10.1080/00401706.1972.10488930>
- Stigler SM (1977) Do robust estimators work with real data? *Ann Stat.* <https://doi.org/10.1214/aos/1176343997>
- Teunissen PJG (1985) *Quality control in geodetic networks*. Springer, Berlin Heidelberg
- Teunissen PJG (2000) *Testing theory: an introduction*. Delft University of Technology, Delft
- Teunissen PJG (2018) Distributional theory for the DIA method. *J Geod* 92:59–80. <https://doi.org/10.1007/s00190-017-1045-7>
- Teunissen PJG (1990) An integrity and quality control procedure for use in multi sensor integration. In: *Proceedings ION GPS* (republished in ION Red Book Series, vol. 7, 2010), pp 15
- Thompson WR (1935) On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. *Ann Math Stat* 6:214–219
- Wiśniewski Z (2009) Estimation of parameters in a split functional model of geodetic observations (M split estimation). *J Geod* 83:105–120. <https://doi.org/10.1007/s00190-008-0241-x>
- Wiśniewski Z (2010) M split(q) estimation: estimation of parameters in a multi split functional model of geodetic observations. *J Geod* 84:355–372. <https://doi.org/10.1007/s00190-010-0373-7>
- Xu P (1987a) A test method for many outliers. *ITC J* 4:314–317
- Xu P (1987b) A F-T method for outliers. *Geom Inf Sci Wuhan Univ* 12:41–46
- Xu P (2005) Sign-constrained robust least squares, subjective breakdown point and the effect of weights of observations on robustness. *J Geod* 79:146–159. <https://doi.org/10.1007/s00190-005-0454-1>
- Yang Y (1994) Robust estimation for dependent observations. *Manuscr Geod* 1:10–17
- Yang Y, Song L, Xu T (2002) Robust estimator for correlated observations based on bifactor equivalent weights. *J Geod* 76:353–358. <https://doi.org/10.1007/s00190-002-0256-7>
- Yang L, Wang J, Knight NL, Shen Y (2013) Outlier separability analysis with a multiple alternative hypotheses test. *J Geod* 87:591–604. <https://doi.org/10.1007/s00190-013-0629-0>
- Yang L, Shen Y, Li B, Rizos C (2021) Simplified algebraic estimation for the quality control of DIA estimator. *J Geod* 95:14. <https://doi.org/10.1007/s00190-020-01454-9>
- Zaminpardaz S, Teunissen PJG (2019) DIA-datasnooping and identifiability. *J Geod* 93:85–101. <https://doi.org/10.1007/s00190-018-1141-3>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.