Estimation of Feeding Composition of Industrial Process Based on Data Reconciliation

Yusi Luan, Mengxuan Jiang, Zhenxiang Feng and Bei Sun *

School of Automation, Central South University, Changsha 410083, China; 184611053@csu.edu.cn (Y.L.); jiangmengxuan@csu.edu.cn (M.J.); fengzxcsu@csu.edu.cn (Z.F.)
* Correspondence: sunbei@csu.edu.cn

Abstract: For an industrial process, the estimation of feeding composition is important for analyzing production status and making control decisions. However, random errors or even gross ones inevitably contaminate the actual measurements. Feeding composition is conventionally obtained via discrete and low-rate artificial testing. To address these problems, a feeding composition estimation approach based on data reconciliation procedure is developed. To improve the variable accuracy, a novel robust M-estimator is first proposed. Then, an iterative robust hierarchical data reconciliation and estimation strategy is applied to estimate the feeding composition. The feasibility and effectiveness of the estimation approach are verified on a fluidized bed roaster. The proposed M-estimator showed better overall performance.

Keywords: data reconciliation; robust estimator; gross error detection; feeding composition

1. Introduction

Complete, accurate and reliable data measurements are important for process modeling, model identification, real-time online optimization and process control. However, the measured values of variables in the actual measurement process are inevitably contaminated by random errors or even gross errors. Therefore, the process measurements will deviate from the real values. In addition, the feeding composition, which is important in making control decisions, cannot be measured online due to the high economic cost and technical limitations. Therefore, it is necessary to process the measurements to guarantee the reliability and observability of the system. Based on this, the estimation of feeding composition can be better performed. Data reconciliation is an effective approach for this aim [1]. The principle of data reconciliation is to minimize the sum of square deviations between the coordinated value of variables and their measurements under the condition of satisfying material balances, heat balances and boundary constraints of process variables. From the mathematical point of view, data reconciliation is a constrained optimization problem [2], which has been widely used in many important procedures such as in-line process monitoring [3,4], enhanced control strategy performance [5,6], parameter estimation [7], soft sensor applications [8], quality control [9] and industrial optimization [10,11] among others [3–11].

Traditional data reconciliation typically assumes that measurements contain only random errors with an average value of zero and normal distributions. Nonetheless, the measurements will be contaminated by gross errors. If gross errors are not processed, the accuracy of data reconciliation results will be reduced. The gross error detection is therefore particularly important for data reconciliation. There are currently three main methods for handling measurements with gross errors. The first strategy is sequential gross error detection and data reconciliation. A variety of statistical tests, including the global test [12], the measurement test [13], the nodal test [14], generalized likelihood ratio test [15] and principal component analysis test [16], are used for gross error detection. Following the detection and elimination of gross errors, the traditional data reconciliation is carried

out. The second strategy is simultaneous gross error detection and data reconciliation. Yu et al. [17] proposed a support vector regression method for recursive simultaneous data reconciliation and gross error detection. Zhang et al. [18] used a novel particle filter (PF) algorithm based on the measurement test (MT) to solve the dynamic simultaneous gross error detection and data reconciliation. Yuan et al. [19] established a new hierarchical Bayesian framework, which can simultaneously estimate the real value of process variables and obtain the magnitudes of gross errors. The last strategy is robust data reconciliation. The estimations close to the true value can be obtained by reducing the influence of gross errors on data reconciliation results. Tjoa and Biegler [20] constructed an M-estimator based on mixed distribution, which takes both random errors and gross errors into account and minimized the M-estimator as an objective function. Johnson and Kramer [21] presented a new robust method for data reconciliation based on a probability bootstrapping theory. Arora and Biegler [22] evaluated the Fair and Hampel estimators in stationary and dynamical constrained problems. Hampel's three-part estimator presented the best performance. Wang and Romagnoli [23] designed a partially adaptive estimator based on a generalized T distribution and a fully adaptive estimator based on density estimation by studying the robustness of a set of existing estimators. Özyurt and Pike [24] analyzed the performance of a number of robust estimators such as Hampel, Cauchy and Fair. Ragot et al. [25] adopted a cost function that has been improved based on the contaminated probability density distribution model for the design of robust estimator. Prata et al. [26] applied the proposed strategy based on the particle swarm optimization (PSO) and the robust Welsch estimator to real-time online detection of Bulk polymerization of propylene. Zhang et al. [27] constructed a robust estimator by using the quasi-weighted least squares. Through the comparative analysis of various methods like Welsch, quasi-weighted least squares and comentropy M-estimators, the feasibility and effectiveness of robust estimators in simultaneous gross error detection and data reconciliation were demonstrated [28]. Alighardashi et al. [29] proposed a maximum likelihood framework for simultaneous data reconciliation and gross error detection for steady-state data. Xie et al. [30] utilized a novel robust estimator to improve the robustness of data reconciliation. Forty-eight different robust M-estimators can be founded in a recent review study, including Fair, Cauchy, Biweight, Jin, Welsch estimator and Xie [31].

The key to estimating feeding composition based on data reconciliation is the robust estimator and the design of the estimation strategy. An excellent robust estimator is embodied in two aspects: On the one hand, the objective function is bounded, and the bound value is small; on the other hand, the influence function can converge quickly and converge to zero. The objective functions of Cauchy [32] and Fair [33] estimators increase with the increase in standard measurement residuals, indicating that the error will greatly affect the reconciled results. Besides, the convergence speed of their influence functions is very slow, and the bound values are large, which shows that the two estimators cannot suppress the influence of gross errors on reconciled results. The objective function of Welsch [26] estimator is limited, but the bound value is also large, and the influence function does not converge rapidly. Although the influence function of Xie [30] estimator may converge to zero and the convergence speed is improved in comparison to the estimators above, it still cannot achieve the fast standard. Based on the robust estimation theory, a novel robust estimator is proposed to enhance the performance of the robust estimators. The objective function is bounded, the bound value is smaller, and the influence function converges toward zero faster. At the same time, aiming at the problem that the change of feeding composition caused by the long test period of industrial processes is unknown, an iterative robust hierarchical data reconciliation and estimation strategy based on the heat balance is proposed. In order to further verify the robustness and effectiveness of the proposed robust estimator and estimation of feeding composition strategy, a series of comparative experiments are carried out through two numerical examples and the real data from a fluidized bed roaster for zinc smelting.

After initial remarks, motivation aspects and applications of the data reconciliation issue, the paper is organized as follows. Section 1 presents preliminary approaches to traditional data reconciliation and robust estimator. The novel robust estimator is proposed, and two numerical examples are used to demonstrate the effectiveness of the proposed estimator in Section 2. Section 3 provides an iterative robust hierarchical data reconciliation and estimation strategy to estimate feeding composition. The feeding composition of fluidized bed roaster is then estimated in Section 4. Finally, Section 5 provides the conclusions.

2. Preliminaries

2.1. Data Reconciliation

The purpose of data reconciliation is to obtain the reconciled data by processing the raw measurements. The reconciled data not only satisfy the constraints of the process model but are also closer to the true value. There are three main assumptions for the system using data reconciliation: The system is in a steady state, the measurement errors follow a normal distribution with zero mean, and each measured variable is independent of each other. Taking into account the presence of measurement errors, the measurement model may be expressed as follows:

$$x = x^* + \varepsilon \tag{1}$$

where x denotes the vector of raw process measurements, x^* denotes the vector of true values of the process variables and ε is the random measurement errors.

Based on the principle of data reconciliation, the general steady-state data reconciliation problem can be stated as a form of solving the weighted least squares solution satisfying the process model and boundary constraints:

$$
\begin{aligned}
&\min \ (x - \widetilde{x})^T \textstyle\sum^{-1}(x - \widetilde{x}) \\
&s.t. \quad F(\widetilde{x}, u) = 0 \\
&\qquad\ \ G(\widetilde{x}, u) \leq 0
\end{aligned}
\tag{2}
$$

where $\widetilde{x}$ is the vector of reconciled data, $\sum$ is the diagonal covariance matrix of measurement errors, u is the vector of unmeasured variables; F represents the process model, which is used as equality constraints in the optimization problem, and G denotes the inequality constraints indicating variable boundaries.

2.2. M-Estimator

In the classical weighted least square of data reconciliation, Equation (2), it is usually assumed that the process measurements contain only random errors. However, in the actual measurement process, the measured values may contain gross errors. The presence of gross errors has a serious effect on the conventional data reconciliation, resulting in the propagation of gross errors. This will cause the reconciled data to not satisfy the process model and deviate from the true value of variables seriously. As a result, the robust estimator is used to account for gross errors in measurements.

Currently, there are numerous approaches for robust data reconciliation, the majority of which are based on the theory of M-estimator. The function of the measurement residuals is constructed as the objective function of the robust estimator. By reducing the weight of the measurements with gross errors, it can prevent gross errors smearing other measurements. Hence, the robust estimator can well reconcile the data with gross errors. The problem of robust data reconciliation can be expressed as follows:

$$
\begin{aligned}
&\min \sum_{i=1}^{n} \rho(\xi_i) = \min \sum_{i=1}^{n} \rho\!\left(\tfrac{x_i - \widetilde{x}_i}{\sigma_i}\right) \\
&s.t. \quad F(\widetilde{x}, u) = 0 \\
&\qquad\ \ \widetilde{x}_{i\min} \leq \widetilde{x}_i \leq \widetilde{x}_{i\max}, \ i = 1, 2, \ldots, n \\
&\qquad\ \ u_{l\min} \leq u_l \leq u_{l\max}, \ l = 1, 2, \ldots, N - n
\end{aligned}
\tag{3}
$$

where ρ is the robust estimator, $\xi_i = (x_i - \widetilde{x}_i)/\sigma_i$ is the standardized residual for the ith measured variable, x_i, $\widetilde{x}_i$ and u_l are, respectively, the measured data, the reconciled data for the ith measured variable and the estimated value for the jth unmeasured variable, $\widetilde{x}_{i\min}$, $\widetilde{x}_{i\max}$, $u_{l\min}$ and $u_{l\max}$ are, respectively, the lower limit, the upper limit for the ith reconciled variable, the lower limit, the upper limit for the lth unmeasured variable. N, n, $N - n$ denote the total number of variables, the total number of measured variables and the total number of variables not measured, respectively. σ_i represents the standard deviation for variable measurement errors.

Most M-estimators are not based on a clearly defined probability distribution known in advance. Most of them are based on a simple mathematical structure. The mathematical structure of M-estimator, i.e., ρ, has some general characteristics as follows:

- ρ is is continuous;
- $\rho(\xi) = \rho(-\xi)$;
- $\rho(\xi) \geq 0$ and ρ is integrable
- $\rho(\xi_1) \leq \rho(\xi_2)$, for $|\xi_i| < |\xi_j|$;
- $\rho(0) = 0$.

As a crucial index in assessing the robustness of the M-estimator, the influence function (IF) [34] is directly proportional to the derivative of the objective function of the estimator. The influence function value refers to the effect of different deviation on the estimator. In general, the function may be defined as the first derivative of the objective function to standardized residuals, which can be expressed as follows:

$$\mathrm{IF}(\xi) = \frac{d\rho(\xi)}{d\xi} \tag{4}$$

Some general characteristics of IF [35] are:

- IF is limited;
- IF is continuous or piecewise continuous;
- $\mathrm{IF}(-\xi) = -\mathrm{IF}(\xi)$;
- IF is nearly linear near the origin ($\mathrm{IF}(\xi) \approx k \cdot \xi, k \neq 0$, for small ξ), but this characteristic is not necessary;
- the rejection point of IF (the point where IF is zero) is finite to suppress large deviations.

On the basis of a robust M-estimation theory, the influence function needs to be continuous and bounded. When ξ is small, the influence function is proportional to ξ; when ξ is infinite, the influence function can converge to a constant, which indicates that the robust estimator can suppress the effect of gross errors in the reconciled results. The influence function of the weighted least squares estimator is the standardized residual ξ, which indicates that the influence function will increase with the measurement errors. It can be seen that the weighted least squares estimator is not robust because its influence function is boundless, and the reconciled results can be easily affected by gross errors.

3. Data Reconciliation Based on a Novel Robust Estimator

3.1. A Novel Robust Estimator

The most robust estimators are constructed according to the following requirements: The objective function of the estimator is bounded, and the bound value is small; the influence function can rapidly converge to a constant. From the observation of the mathematical structure of 48 different robust estimators in the literature [31], it has been shown that the objective functions of these M-estimators comprise limitless function, limited function, non-piecewise function, piecewise function, multi parameters and a single parameter. The performance of the M-estimator whose objective function is limited, non-piecewise and has single parameter is analyzed. It is found that the M-estimator with a relatively rapid convergence rate basically contains the mathematical expression with e as the base number. The derivative of exponential function based on e is equal to itself. Moreover, $e^{-\xi}$ decreases with increasing ξ and tends towards 0. When ξ tends towards 0, $e^{-\xi}$ tends towards 1.

Therefore, the objective function can be limited if the structure of $1 - e^{-\xi}$ is included in. For instance, the Welsch estimator and the objective function is presented in Equation (5). In order to improve the robustness of the M-estimator, Xie divided $1 - e^{-\xi}$ with $1 + e^{-\xi}$. By the decline of $1 + e^{-\xi}$, the overall rise speed of the objective function is improved. Furthermore, the converging speed of the influence function is accelerated. The objective function of Xie estimator is illustrated under Equation (6).

Welsch:

$$\rho(\xi_i) = \frac{c_w^2}{2}\left(1 - \exp\left(-\left(\frac{\xi_i}{c_w}\right)^2\right)\right) \tag{5}$$

Xie:

$$\rho(\xi_i) = \frac{1 - \exp\left(-\left(\frac{\xi_i}{c_x}\right)^2\right)}{1 + \exp\left(-\left(\frac{\xi_i}{c_x}\right)^2\right)} \tag{6}$$

where c_w and c_x are, respectively, tuning parameters of Welsch and Xie estimators.

As discussed above, in order to limit the effect of gross errors on the reconciled results, a novel robust estimator is created based on the general characteristics of the M-estimator and influence function. The objective function of the novel robust estimator is defined as follows:

$$\rho(\xi_i) = \frac{c_p^2\left[1 - \exp\left(-(\xi_i/c_p)^2\right)\right]}{4\left[1 + \exp\left(-(\xi_i/c_p)^4\right)\right]} \tag{7}$$

where c_p is the tuning parameter; the definition of ξ_i is the same as above. By increasing the index of ξ, the downward velocity of $1 + e^{-\xi}$ is increased. Then, the objective function increases faster, and the influence function converts to 0 faster. At the same time, in order to satisfy the characteristic that the influence function is almost linear near the origin, the objective function is multiplied by $c_p^2/4$. The influence function of the novel robust estimator function expressed in Equation (7) is as follows:

$$\mathrm{IF}(\xi_i) = \frac{d\rho(\xi_i)}{d\xi_i} = \frac{\xi_i \exp\left(-\left(\frac{\xi_i}{c_p}\right)^2\right)\left[1 + \exp\left(-\left(\frac{\xi_i}{c_p}\right)^4\right) + \frac{2\xi_i^2}{c_p^2}\left(1 - \exp\left(-\left(\frac{\xi_i}{c_p}\right)^2\right)\right)\right]}{2\left[1 + \exp\left(-\left(\frac{\xi_i}{c_p}\right)^4\right)\right]^2} \tag{8}$$

It can be seen from Equations (7) and (8) that when the standardized residual is small, the objective function of the novel robust estimator tends to be gradually zero, and the influence function is proportional to the standardized residual. It shows that the weighted least squares method can be used to reconcile the data where measurements contain only random errors. When the standardized residual is significant, the objective function of the novel robust estimator tends to be constant little by little, and the influence function quickly converges to zero. It is demonstrated that the novel robust estimator can adjust the weight of the estimator based on the standardized residual size when the measurements contain gross errors. The larger the gross errors, the lower the weights, and the greater the effects of restraining the gross errors. Therefore, the proposed estimator is more robust.

3.2. The Tuning Parameter of the Novel Robust Estimator

The use of robust estimators in data regression typically involves the inverse relationship between relative efficiency and robustness [36]. Relative efficiency refers to the fit quality of the estimated value reconciled by M-estimator, when errors follow another distribution relative to the reference distribution (usually assumed to be Normal distribution). The so-called robustness refers to the performance of the M-estimator under various nonnormal error distributions. The more robust an M-estimator is, the less its relative efficiency is. The relationship between robustness and relative efficiency is addressed by the tuning parameters. When comparing the performance of different M-estimators, the

level of relative efficiency and the reference distribution should be consistent. In general, the tuning parameters for each M-estimator are obtained at the relative efficiency level of 95%. Then the robustness of distinct M-estimators is judged.

The mathematical definition of relative efficiency is illustrated in Equation (9):

$$E_{ff}[\text{IF}(\xi), f(\xi)] = \frac{V_f[\text{IF}_f(\xi), f(\xi)]}{V[\text{IF}(\xi), f(\xi)]} \tag{9}$$

where V_f is the asymptotic variance of the reference estimator, V is the asymptotic variance of the specified estimator, $f(\xi)$ is the error reference probability distribution, $\text{IF}_f(\xi)$ is the influence function of the reference estimator and $\text{IF}(\xi)$ is the influence function of the specified estimator. The asymptotic variance V is defined as follows [37]:

$$V[\text{IF}(\xi), f(\xi)] = \frac{\int_{-\infty}^{+\infty} \text{IF}^2(\xi) f(\xi) d\xi}{\left[\int_{-\infty}^{+\infty} \text{IF}'(\xi) f(\xi) d\xi\right]^2} \tag{10}$$

where $\text{IF}'(\xi)$ is the derivative of the influence function $\text{IF}(\xi)$. Since $\text{IF}'(\xi)$ may be discontinuous, Equation (10) can be further expressed by Equation (11):

$$V[\text{IF}(\xi), f(\xi)] = \frac{\int_{-\infty}^{+\infty} \text{IF}^2(\xi) f(\xi) d\xi}{\left[-\int_{-\infty}^{+\infty} \text{IF}(\xi) f'(\xi) d\xi\right]^2} = \frac{\int_{-\infty}^{+\infty} \text{IF}^2(\xi) f(\xi) d\xi}{\left[-\int_{-\infty}^{+\infty} \text{IF}(\xi) f'(\xi) d\xi\right]^2} \tag{11}$$

where $f'(\xi)$ is the derivative of $f(\xi)$.

The tuning parameter of the proposed estimator at the relative efficiency level of 95% in respect to the Normal distribution is calculated as follows:

(1) The Normal distribution is selected as the reference distribution, and its probabilistic distribution and the first derivative are respectively expressed as follows:

$$f(\xi) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2}\right) \tag{12}$$

$$f'(\xi) = -\xi \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2}\right) = -\xi f(\xi) \tag{13}$$

(2) The least square estimator is selected as the reference estimator, and its influence function is expressed as follows:

$$\text{IF}_f(\xi) = \frac{d\rho(\xi)}{d\xi} = \frac{d}{d\xi}\left(\frac{\xi^2}{2}\right) = \xi \tag{14}$$

(3) The influence function of the proposed estimator is shown in Equation (8). The expressions of V_f and V can be obtained by using Equation (11):

$$V_f = \frac{2 \cdot \int_0^{+\infty} \text{IF}_f^2(\xi) f(\xi) d\xi}{\left(2 \cdot \int_0^{+\infty} \text{IF}_f(\xi) \xi f(\xi) d\xi\right)^2} \tag{15}$$

$$V = \frac{2 \cdot \int_0^{+\infty} \text{IF}^2(\xi) f(\xi) d\xi}{\left(2 \cdot \int_0^{+\infty} \text{IF}(\xi) \xi f(\xi) d\xi\right)^2} \tag{16}$$

(4) The expression of relative efficiency can be obtained from Equation (9). Making the expression equal to 0.95, a univariate equation on the tuning parameter c_p can be

constructed, as shown in Equation (17). By numerical calculation, the tuning parameter at the relative efficiency level of 95% in respect to the Normal distribution is 1.5424.

$$\frac{V_f}{V} = \frac{\dfrac{2\int_0^{+\infty} \xi^2 \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp(-(\xi^2/2))d\xi}{\left[2\int_0^{+\infty} \xi^2 \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp(-(\xi^2/2))d\xi\right]^2}}{\dfrac{2\int_0^{+\infty} \frac{\xi^2 \exp(-2(\xi/c_p)^2)[1+\exp(-(\xi/c_p)^4)+2\xi^2(1-\exp(-(\xi/c_p)^2))/c_p^2]^2}{4[1+\exp(-(\xi/c_p)^4)]^4} \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp(-(\xi^2/2))d\xi}{\left[2\int_0^{+\infty} \frac{\xi^2 \exp(-(\xi/c_p)^2)[1+\exp(-(\xi/c_p)^4)+2\xi^2(1-\exp(-(\xi/c_p)^2))/c_p^2]}{2[1+\exp(-(\xi/c_p)^4)]^2} \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp(-(\xi^2/2))d\xi\right]^2}} = 0.95 \quad (17)$$

To more accurately reflect the competitive relationship between the relative efficiency and robustness of the M-estimator, the comparative images of the objective function and influence function of the proposed, Xie and Welsch estimators are shown in Figure 1. Relative efficiency levels are 90%, 95%, 98% and 99%, respectively. The tuning parameters of the three estimators at different relative efficiency levels are presented in Table 1. Table 1 and Figure 1 show that the higher the relative efficiency level, the higher the values of the tuning parameters for the three estimators. The higher the tuning parameter, the slower the convergence speed of the influence function to 0 and the weaker the robustness. Hence, when comparing the robustness of distinct estimators, the comparative analysis has to be performed at the same level of relative efficiency.

Table 1. Tuning parameters of three estimators at distinct relative efficiency levels.

	M-Estimator	Tuning Parameter	
1	Welsch	$c_w = 2.3828$	$E_{ff} = 90\%$
		$c_w = 2.9846$	$E_{ff} = 95\%$
		$c_w = 3.9077$	$E_{ff} = 98\%$
		$c_w = 4.7343$	$E_{ff} = 99\%$
2	Xie	$c_x = 1.6705$	$E_{ff} = 90\%$
		$c_x = 1.9597$	$E_{ff} = 95\%$
		$c_x = 2.3409$	$E_{ff} = 98\%$
		$c_x = 2.6359$	$E_{ff} = 99\%$
3	Proposed	$c_p = 1.3082$	$E_{ff} = 90\%$
		$c_p = 1.5424$	$E_{ff} = 95\%$
		$c_p = 2.0942$	$E_{ff} = 98\%$
		$c_p = 2.6060$	$E_{ff} = 99\%$

In order to further analyze the effectiveness of the proposed estimator, the objective function and influence function of the proposed estimator are compared with the other four robust estimators. The four types of robust estimators are Fair, Cauchy, Welsch and Xie. The objective functions of Fair and Cauchy are shown in Equations (18) and (19):
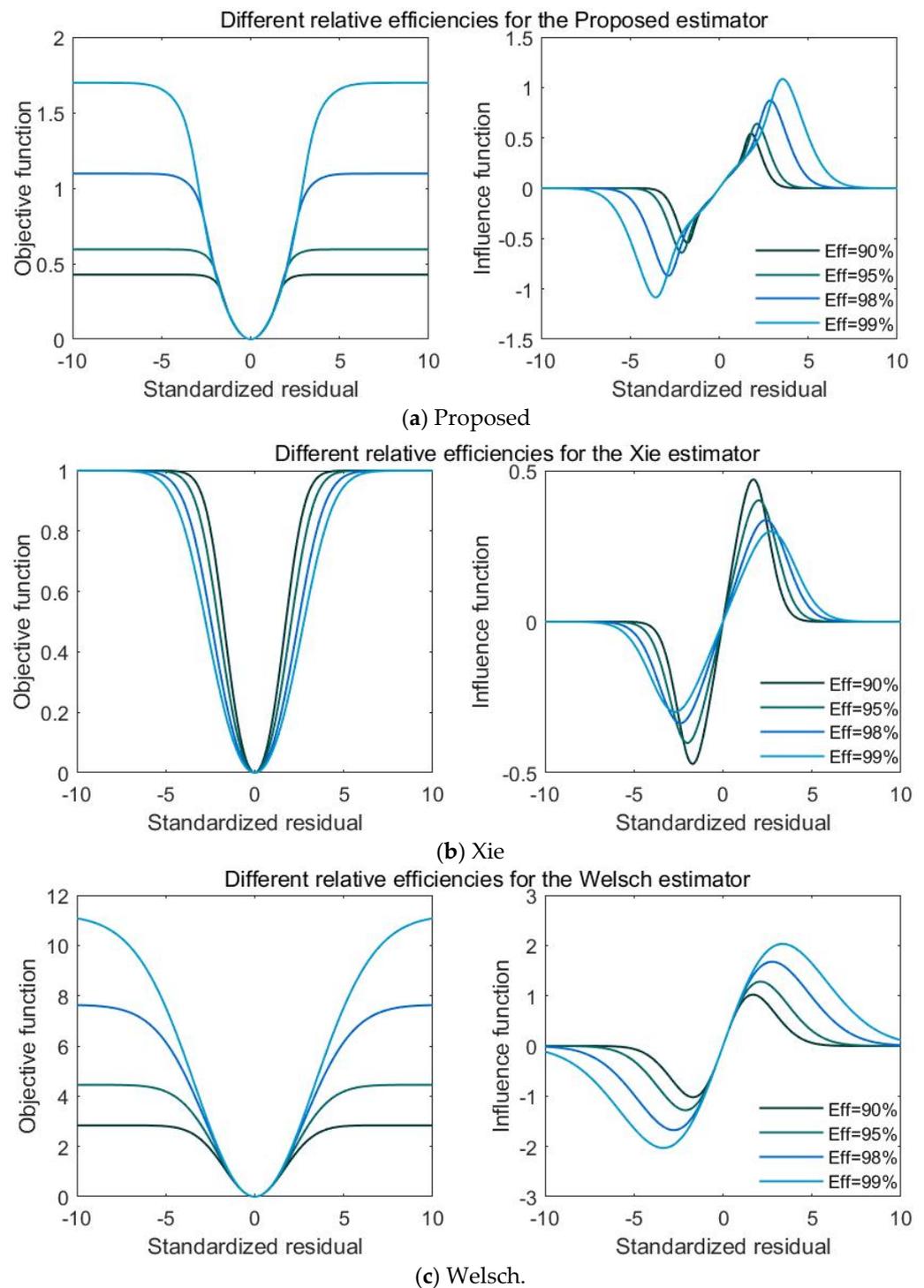
Fair:

$$\rho(\xi_i) = c_f^2 \left(\frac{|\xi_i|}{c_f} - \ln\left(1 + \frac{|\xi_i|}{c_f}\right) \right) \quad (18)$$

Cauchy:

$$\rho(\xi_i) = c_c^2 \ln\left(1 + \frac{\xi_i^2}{c_c^2}\right) \quad (19)$$

where c_f and c_c are, respectively, tuning parameters of Fair and Cauchy estimators. The tuning parameters of the five M-estimators under the level of 95% relative efficiency are $c_f = 1.3998$, $c_w = 2.9846$, $c_c = 2.3849$, $c_x = 1.9597$ and $c_p = 1.5424$. The results of the comparison are illustrated in Figures 2 and 3.

Figure 1. Characteristicfunctions of three estimators at distinct relative efficiency levels.

As can be seen from Figures 2 and 3, the objective function and influence function of the proposed M-estimator satisfy their general characteristics. When the standardized residual is low, the objective and influence functions of several robust estimators are all close, which indicates that low measurement errors have little effect on these estimators. When the standardized residual is large, the objective function of the Fair and Cauchy estimator increases at a higher rate. Their influence functions do not converge rapidly and eventually converge to a nonzero value. It shows that both Fair and Cauchy estimators are sensitive to gross errors. Although the objective function of Welsch does not diverge, it tends to be a greater constant than Xie and the novel robust estimator. The objective

functions of the novel robust estimator and Xie estimator are less than those of the other three estimators in the case of gross errors in measurements. Furthermore, the proposed M-estimator tends to be a constant lower than Xie. It can be seen further from Figure 2 that when there are gross errors in measurements, the influence function of the proposed M-estimator decreases much faster than that of Xie. When the standardized residual exceeds 4, the influence function of the proposed M-estimator converges to 0. Nevertheless, the influence function of Xie converges to 0 only when the standardized residual is above 5. Hence one can see that the convergence speed of influence function of the novel robust estimator is faster than that of Xie, and the effect of suppressing gross errors is excellent. The proposed method is therefore more robust and efficient than other estimators.
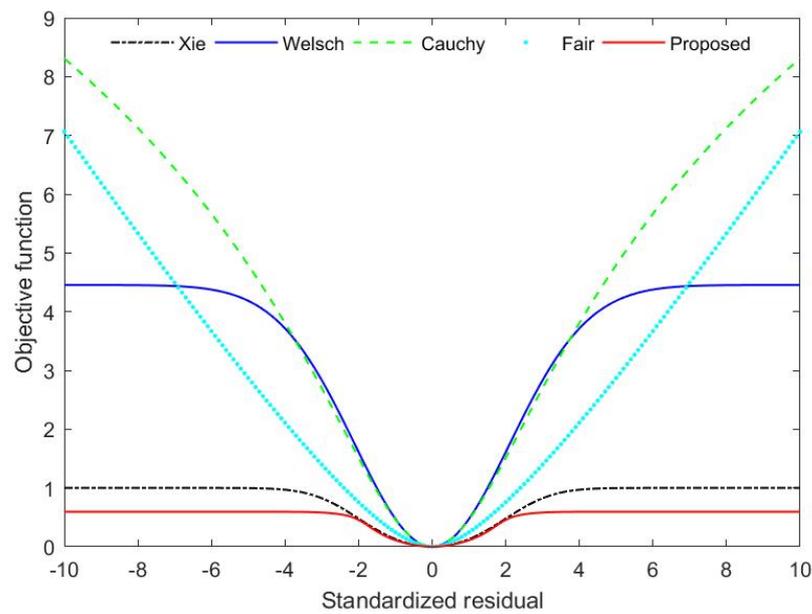

Figure 2. Objective functions for Xie, Welsch, Cauchy, Fair and Proposed estimator.
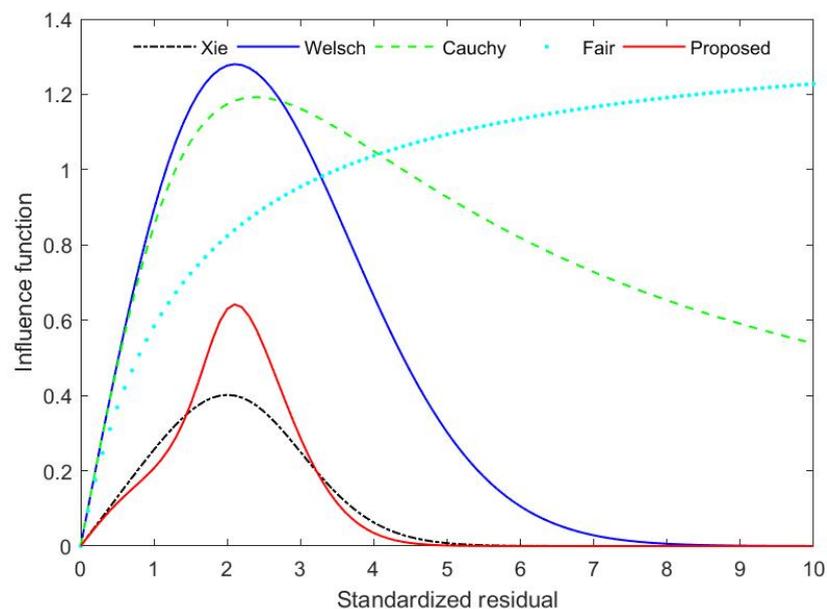

Figure 3. Influence functions for Xie, Welsch, Cauchy, Fair and Proposed estimator.

Next, two numerical examples are used to further verify the effectiveness of the proposed robust estimator, and the proposed robust estimator is compared with several

estimators. The data reconciliation procedure can be interpreted mathematically as an optimization problem. In this paper, the state transition algorithm (STA) is applied as the optimization method for data reconciliation problems [38]. The local search operator, the global search operator and the heuristic search operator are adopted as the state transformation operators. This can prevent the solution process from falling into the local optimum and reduce the search time for the global optimum.

3.3. Linear Case

In this part, the measurement network [26] as shown in Figure 4 is adopted. The network is a linear structure consisting of four nodes and seven streams. As can be seen from Figure 4, the seven streams meet the following material balance

$$
\begin{aligned}
x_1 - x_2 + x_4 + x_6 &= 0 \\
x_2 - x_3 &= 0 \\
x_3 - x_4 - x_5 &= 0 \\
x_5 - x_6 - x_7 &= 0
\end{aligned}
\tag{20}
$$

where $x = [x_1, x_2, \ldots, x_7]^T$ is set of stream variables. In this linear example, all variables are assumed to be measured, and the true value of each variable is X = [5, 15, 15, 5, 10, 5, 5]. The true value of each variable is added with some random errors to get the corresponding measurements. Let the standard deviation of random errors be 2.5% of the true value of each stream. The diagonal matrix of the variance is as follows:

$$
\Sigma = 0.01 \times diag(1.562, 14.062, 14.062, 1.562, 6.250, 1.562, 1.562)
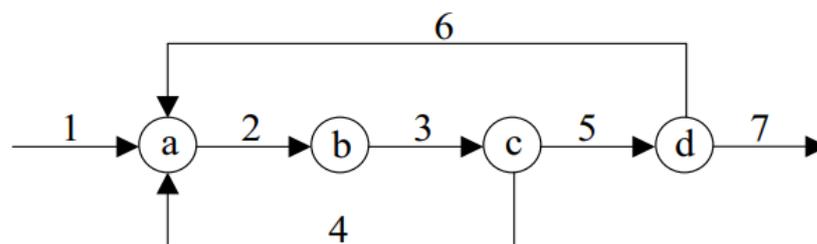\tag{21}
$$

Figure 4. Diagram of the measurement network.

For a precise analysis of the performance of each robust estimator, SSE (sum of squares due to error), TER (total error reduction) and RER (relative error reduction) [27] are used as indicators to assess the precision of the reconciled results. The definitions included are:

$$
\text{SSE} = \sum_{i=1}^{n} (\tilde{x}_i - x_i^*)^2 + \sum_{j=1}^{m} (\tilde{u}_j - u_j^*)^2
\tag{22}
$$

$$
\text{TER} = \frac{\sqrt{\sum_{i=1}^{n} \left(\frac{x_i - x_i^*}{\sigma_i}\right)^2} - \sqrt{\sum_{i=1}^{n} \left(\frac{\tilde{x}_i - x_i^*}{\sigma_i}\right)^2}}{\sqrt{\sum_{i=1}^{n} \left(\frac{x_i - x_i^*}{\sigma_i}\right)^2}}
\tag{23}
$$

$$
\text{RER} = \frac{\sum_{i=1}^{n} \left(\frac{|x_i^* - x_i|}{x_i^*} - \frac{|x_i^* - \tilde{x}_i|}{x_i^*}\right)}{\sum_{i=1}^{n} \frac{|x_i^* - x_i|}{x_i^*}}
\tag{24}
$$

where $\tilde{x}_i$ represents the reconciled data for the ith measurable variable; x_i^* represents the true value for the ith measurable variable; x_i represents the measured data for the ith measurable variable; $\tilde{u}_j$ is the estimated value of the jth unmeasurable variable; u_j^* is the

true value of the jth unmeasurable variable; σ_i denotes the standard deviation for the ith measurable variable; n is the total number of measurable variables; m is the total number of unmeasurable variables. The smaller the SSE, the closer the reconciled results are to the true value. The larger the TER, the more accurate the reconciled results. The larger RER shows that the reconciled results are less affected by gross errors.

3.3.1. There Are Two Gross Errors in the Measurement Variables

Streams 2 and 5 are selected to add gross errors in the magnitude of 2 and 1, respectively, and other streams contain only random errors. The results of the data reconciliation and indicators for several methods are presented in Table 2. For convenience, "Proposed" is used to represent the proposed estimator in this table and subsequent tables.

Table 2. Reconciled results of different methods with two gross errors.

Stream	True	Original Meas.	Meas. with Gross Error	Proposed	Xie	Welsch	Cauchy	Fair
x_1	5	4.995	4.995	5.0111	5.0152	5.0346	5.0519	5.1777
x_2	15	14.91	16.91	15.0420	15.0540	15.1202	15.1970	15.6413
x_3	15	15.01	15.01	15.0420	15.0540	15.1202	15.1970	15.6413
x_4	5	5.002	5.002	4.9987	4.9985	5.0053	5.0230	5.1795
x_5	10	9.98	10.98	10.0433	10.0555	10.1149	10.1740	10.4618
x_6	5	5.019	5.019	5.0322	5.0403	5.0803	5.1221	5.2841
x_7	5	5.014	5.014	5.0111	5.0152	5.0346	5.0519	5.1777
SSE	–	–	–	0.0067	0.0110	0.0510	0.1287	1.2117
TER	–	–	–	0.9424	0.9263	0.8457	0.7606	0.2953
RER	–	–	–	0.9101	0.8838	0.7501	0.6008	−0.2626

Table 2 shows that the SSE of the proposed robust estimator (Proposed) is much lower than that of Fair with 1.2117 and Cauchy with 0.1287 and is less than that of Xie and Welsch estimator. The TER of the novel robust estimator is much larger than that of Fair with 0.2953 and Cauchy with 0.7606, and higher than that of Xie and Welsch estimator. In addition, the RER of Proposed is also higher than that of the other four robust estimators. The results show that in the case of two gross errors, the reconciled results of the novel robust estimator are closer to the true value, and the proposed estimator performs better than other methods.

3.3.2. There Are Three Gross Errors in the Measurement Variables

Streams 2, 5 and 7 are selected to add gross errors with the magnitudes of 2, 1 and −1.5, respectively, and other streams contain only random errors. The results of the data reconciliation and indicators for several methods are presented in Table 3.

Table 3. Reconciled results of different methods with three gross errors.

Stream	True	Original Meas.	Meas. with Gross Error	Proposed	Xie	Welsch	Cauchy	Fair
x_1	5	4.995	4.995	5.0080	5.0164	5.0574	5.0323	4.3533
x_2	15	14.91	16.91	15.0390	15.0552	15.1425	15.1802	15.1833
x_3	15	15.01	15.01	15.0390	15.0552	15.1425	15.1802	15.1833
x_4	5	5.002	5.002	4.9991	4.9984	5.0037	5.0244	5.3066
x_5	10	9.98	10.98	10.0399	10.0568	10.1388	10.1558	9.8767
x_6	5	5.019	5.019	5.0320	5.0404	5.0814	5.1235	5.5233
x_7	5	5.014	3.514	5.0080	5.0164	5.0574	5.0323	4.3533
SSE	–	–	–	0.0058	0.0115	0.0731	0.1071	1.2866
TER	–	–	–	0.9743	0.9642	0.9111	0.8954	0.3474
RER	–	–	–	0.9641	0.9470	0.8621	0.8446	0.1268

Table 3 shows that the SSE of the proposed robust estimator (Proposed) is much lower than that of Fair with 1.2866 and Cauchy with 0.1071 and is less than that of Xie and Welsch

estimator. The TER of the novel robust estimator is much larger than that of Fair with 0.3474, and higher than that of Xie, Welsch and Cauchy estimator. In addition, the RER of Proposed is also higher than that of the other four robust estimators. The results show that in the case of three gross errors, the reconciled results of the novel robust estimator are closer to the true value, and the proposed estimator performs better than other methods.

In order to better illustrate the effectiveness of the proposed estimator, the linear numerical example is divided into three groups for comparative experiments. Each group comprises 100 data samples. In the first group, there is gross error in stream 2 (x_2). In group two, there are gross errors in stream 2 (x_2) and stream 5 (x_5). The third group contains gross errors in stream 2 (x_2), stream 5 (x_5) and stream 7 (x_7). In each group of experiments, the selected variables are added with gross errors ranging from 3 times to 9 times the standard deviation. The measurement test (MT) is used to detect gross errors. Two criteria are selected to assess the performance of different methods, including the observed power (*OP*) and the Average Number of Type I errors (*ATVI*) [2]. *OP* and *ATVI* are defined as follows:

$$OP = \frac{\text{The number of gross errors detected correctly}}{\text{The total number of gross errors}} \tag{25}$$

$$ATVI = \frac{\text{The number of gross errors detected mistakenly}}{\text{The total number of simulations}} \tag{26}$$

OP indicates the number of gross errors correctly detected. *ATVI* indicates the number of gross errors mistakenly detected. The larger value of the *OP* is, the stronger ability of the estimator to identify gross errors. The lower value of the *ATVI* is, the less frequently gross errors are detected incorrectly. The statistical graph of the process in which gross errors are detected by distinct methods in each group is shown in Figure 5. The abscissa of each graph corresponds to seven stream variables. The y-axis is the number of times each variable is detected to contain gross errors. The *OP* and *ATVI* of the proposed estimator, Xie, Welsch, Cauchy and Fair are obtained by statistical analysis of the process diagram, as illustrated in Figure 6.

As depicted in Figure 6, when one variable, two variables and three variables contain gross errors, the *OP* obtained by the proposed estimator is higher than that obtained by other estimators. The result shows that the proposed estimator has a higher probability of identifying gross errors correctly. In addition, the *AVTI* obtained by the proposed estimator is lower than that obtained by other methods. It can also be seen from Figure 5 that the proposed estimator has a lower number of gross errors mistakenly detected than other estimators. It is shown that gross errors have less influence on the data reconciliation based on the Proposed estimator.

3.4. Nonlinear Case

A nonlinear numerical example [22] is used in this part. Within the nonlinear system, there are five measurable variables, three unmeasurable variables and six nonlinear constraint equations, which are described as Equation (27):

$$\begin{cases} 0.5x_1^2 - 0.7x_2 + x_3u_1 + x_2^2u_1u_2 + 2x_3u_3^2 - 255.8 = 0 \\ x_1 - 2x_2 + 3x_1x_3 - 2x_2u_1 - x_2u_2u_3 + 111.2 = 0 \\ x_3u_1 - x_1 + 3x_2 + x_1u_2 - x_3\sqrt{u_3} - 33.57 = 0 \\ x_4 - x_1 - x_3^2 + u_2 + 3u_3 = 0 \\ x_5 - 2x_3u_2u_3 = 0 \\ 2x_1 + x_2x_3u_1 + u_2 - u_3 - 126.6 = 0 \end{cases} \tag{27}$$

where $x_i(i = 1, 2, \ldots, 5)$ is the measured variable and $u_j(j = 1, 2, 3)$ is the unmeasured variable. Measurements of the measured variables are obtained from the true value added, some random errors and gross errors. The true value, the measured value, the standard deviation of the measured variables and the true value of the unmeasured variables are

given in Table 4. Amongst these, x_2, x_3 and x_5 contain gross errors. The results of the data reconciliation and indicators for several methods are presented in Table 4.
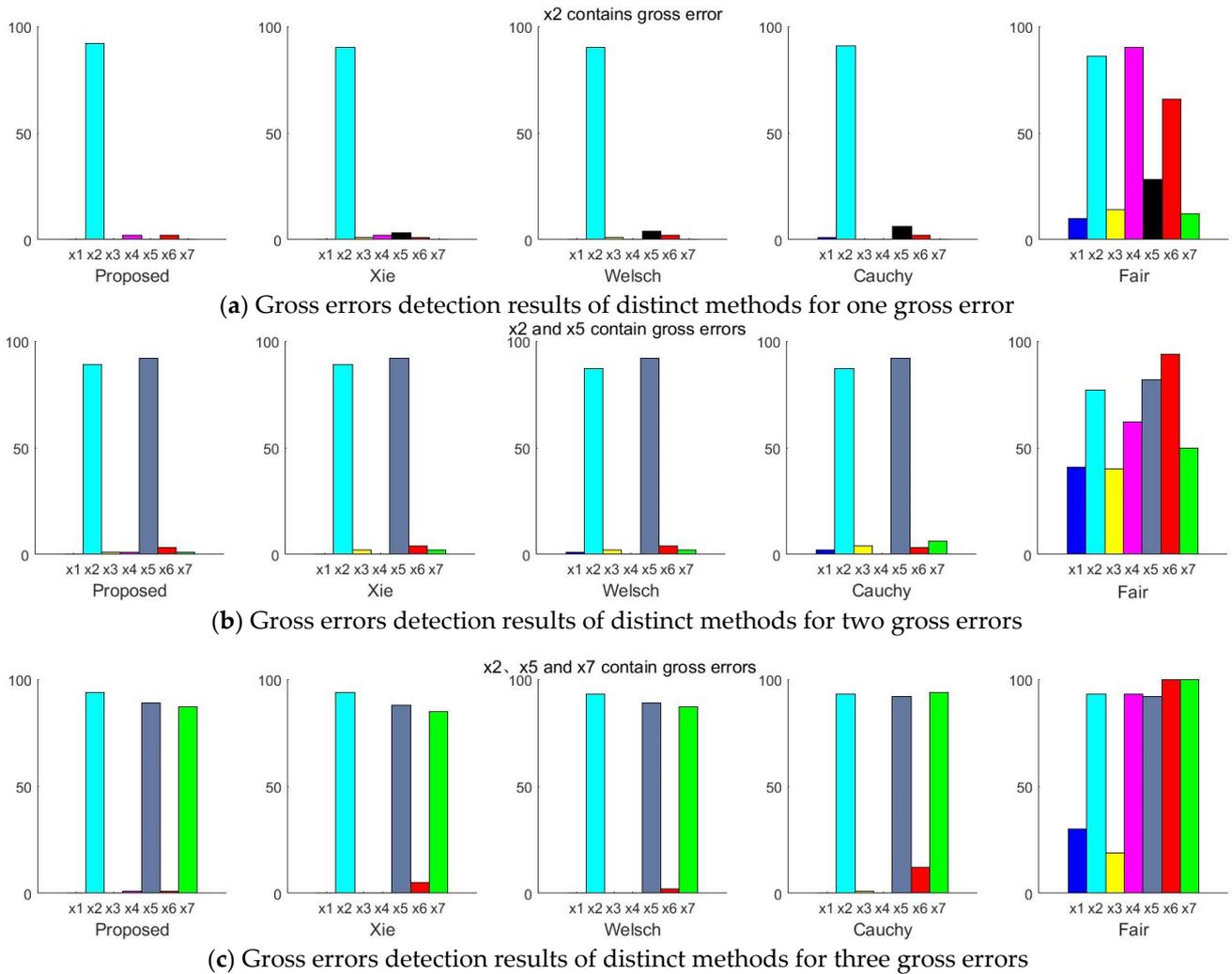

(**a**) Gross errors detection results of distinct methods for one gross error

(**b**) Gross errors detection results of distinct methods for two gross errors

(**c**) Gross errors detection results of distinct methods for three gross errors

Figure 5. Statistical diagram of gross errors detection process for distinct methods.
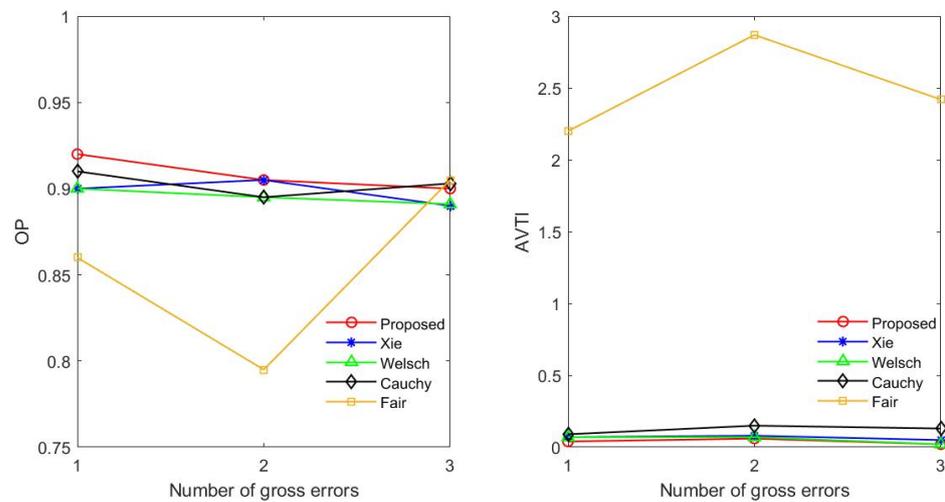
Figure 6. Results of *OP* and *AVTI* with different methods.

Table 4. Reconciled results of different methods for nonlinear constraints.

Stream	Standard Deviation	True	Meas.	Proposed	Xie	Welsch	Cauchy	Fair
x_1	0.5	4.5124	4.5360	4.5378	4.5280	4.5379	4.5796	4.4727
x_2	0.6	5.5819	5.9070	5.5754	5.5655	5.5331	5.5360	5.6514
x_3	0.2	1.9260	1.8074	1.9221	1.9223	1.9200	1.9153	1.9321
x_4	0.2	1.4560	1.4653	1.4653	1.4842	1.4924	1.5096	1.4655
x_5	0.5	4.8545	4.5491	4.8156	4.8010	4.8083	4.7882	4.7870
u_1	–	11.070	–	11.0988	11.1178	11.1939	11.2079	10.9076
u_2	–	0.6147	–	0.6143	0.6160	0.6187	0.6168	0.6104
u_3	–	2.0504	–	2.0469	2.0349	2.0317	2.0345	2.0372
SSE	–	–	–	0.0031	0.0067	0.0222	0.0333	0.0377
TER	–	–	–	0.8950	0.8192	0.7751	0.6631	0.7994
RER	–	–	–	0.8805	0.8008	0.7324	0.5930	0.7693

Table 4 shows that the SSE calculated by the proposed robust estimator is significantly less than that calculated by Fair, Cauchy and Welsch estimator, and it is less than that calculated by Xie estimator. Furthermore, the TER and RER of the proposed estimator are larger than those of the other four estimators. The results show tha for the nonlinear case, the reconciled results of the new robust estimator are also near the true value. Consequently, the proposed robust estimator can effectively suppress the effect of gross errors on the reconciled results and has greater robustness and excellent effectiveness.

4. Feeding Composition Estimation Based on Iterative Data Reconciliation

In order to address the problem caused by the unknown change in feeding composition, an iterative robust hierarchical data reconciliation and estimation strategy is proposed. The method is applied to a fluidized bed roaster for zinc smelting. Firstly, the industrial process of fluidized bed roaster is introduced. At the same time, difficulties in reconciling the fluidized bed roaster are discussed in Section 3.1. Then, on the basis of the establishment of the mechanism balance model of the fluidized bed roaster, specific solutions are presented in accordance with the existing difficulties.

4.1. Industrial Process Description

The fluidized bed roaster is a type of thermal equipment that applies fluidization technology to make materials desulphurized by oxidation roasting in the metallurgical industry. The roasting process of the fluidized bed roaster is a gas-solid reaction procedure. By sending a great deal of air into the oven from the bottom, a very intense exothermic reaction occurs in the material layer of the sulfur ore under the agitation of the air. Oxygen combines with sulfide to form sulfur dioxide and precious metals are converted into oxides or sulphates. The structure diagram for the fluidized bed roaster is shown in Figure 7.

In the fluidized roasting process, accurate and reliable measurements are the basis of process control, functional analysis and production management. However, in the real process, due to inaccurate sensors, equipment leaks and measurement bias, real-time measurements are inevitably affected by random errors and gross errors. That has a serious effect on the accuracy of measurements. In addition, some important variables, such as smoke volume and calcine quality, cannot be measured due to limitations in measurement technology and the environment.

The ZnS content of the feed during the roasting of the zinc concentrate will affect the indices of operation, such as the resistance to roasting, the temperature of the boiling layer and the content in the gaseous effluents. Then, the technical-economic index of the level of zinc soluble in calcine is affected, which field staff can judge from according to roasting. The compositions of mixed zinc concentrate and calcine are sampled only twice a day, once in the morning and once at night. Then the percentage of Zn, S and other elements in the mixed zinc concentrate and that of soluble zinc and soluble sulfur in the calcine are detected by the laboratory. Nevertheless, the sampling frequency of the variables measured

in real time, like feed volume and blast volume, is 5 seconds, which is much greater than that of the laboratory data. Hence, due to the long test cycle of the mixed zinc concentrate composition in the fluidized bed roaster, the laboratory data cannot reflect the change in the real-time feeding composition.
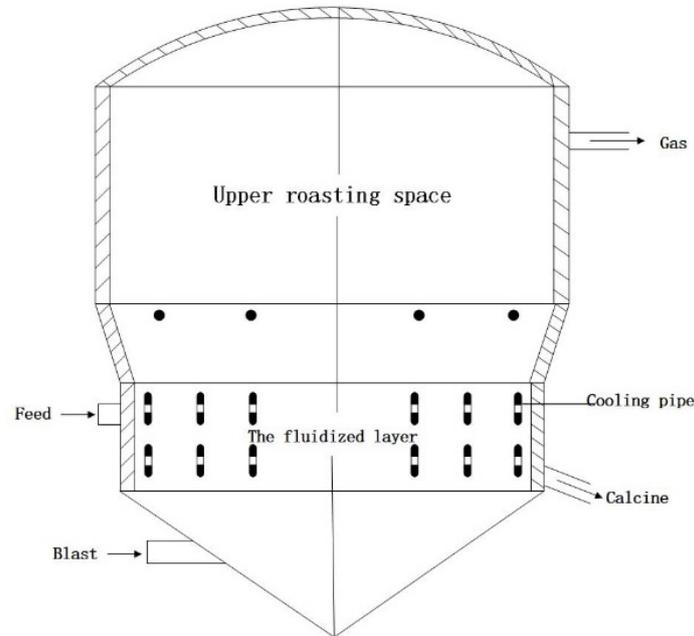

Figure 7. Structure diagram of fluidized bed roaster.

Concerning the problems of measurements contaminated by gross errors, irredundant process data and a long test cycle of the feeding composition, an iterative robust hierarchical data reconciliation and estimation strategy for feeding composition is proposed.

4.2. Iterative Robust Hierarchical Data Reconciliation and Composition Estimation Framework

Figure 8 illustrates the framework of iterative robust hierarchical data reconciliation and estimation strategy for feeding composition. The major steps are as follows:
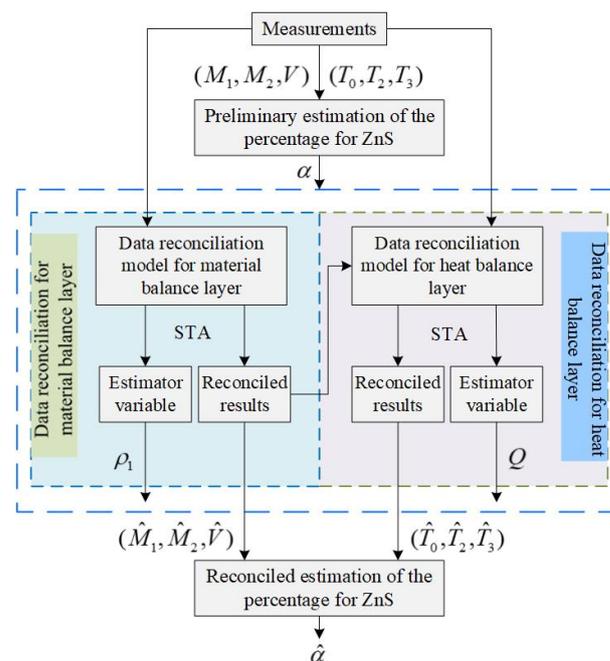
Figure 8. The framework of iterative robust hierarchical data reconciliation and composition estimation.

Step 1: Establish the steady-state mechanism model of the fluidized bed roaster The establishment of a reasonable process model is the basis for data reconciliation. However, the actual industrial process is complex and needs to be simplified accordingly. The following assumptions are put forward to establish a credible and comprehensible mechanism model:

(1) The only exothermal component of the mixed zinc concentrate is ZnS.

(2) Since the specific heat of calcine and soot is close and the mass ratio is close to 1:1, soot can be classified as calcine in the calculation.

Therefore, the material and heat balance at steady-state are expressed by Equations (28) and (29):

Material balance:

$$\frac{16}{97}\alpha(M_1 + M_2)1000 + V_0(\rho_0 - \rho_1) = 0 \tag{28}$$

Heat balance:

$$1000(M_1 + M_2)T_1 cp_1 + V_0 T_0 cp_2 + \frac{443508}{97}\alpha(M_1 + M_2)1000$$
$$= 1000(1 - \frac{16}{97}\alpha)(M_1 + M_2)T_3 cp_3 + V_0 T_2 cp_0 + Q \tag{29}$$

where α denotes the percentage of ZnS in zinc concentrate; M_1 and M_2 represent the first and second feeding quantity, respectively; V_0 is the blast velocity; T_0, T_1, T_2 and T_3 denote the temperature variables for blast, zinc concentrate, gas and calcine, respectively; cp_0, cp_1, cp_2 and cp_3 are specific heat for gas, zinc concentrate, blast and calcine, respectively; Q describe the heat loss; ρ_0 and ρ_1 are, respectively, the density of air and gas. Among these variables, M_1, M_2, V_0, T_0, T_2 and T_3 are measured variables. ρ_1, Q and α are unmeasured variables. The other parameters are all fixed values.

Step 2: Preliminary estimation of α based on measurements and heat balance. Estimation of α can be considered as an optimization problem. Combined with the heat balance and measurements, the ZnS estimation model before data reconciliation could be expressed as follows:

$$\min Y_1 = (1000(M_1 + M_2)T_1 cp_1 + V_0 T_0 cp_2 + \frac{443,508}{97}\alpha(M_1 + M_2)1000$$
$$- 1000(1 - \frac{16}{97}\alpha)(M_1 + M_2)T_3 cp_3 - V_0 T_2 cp_0 - Q)^2 \tag{30}$$

$$s.t. \quad \alpha_{\min} \le \alpha \le \alpha_{\max} \tag{31}$$

where Y_1 is the objective of the optimization problem; $\alpha_{\min}$ and $\alpha_{\max}$ are the top and bottom bounds of α, respectively. In this case, Q can be roughly considered as a fixed value, which can be obtained by expert knowledge. The preliminary estimation of α will be used to construct the hierarchical model in Step 3.

Step 3: Robust hierarchical data reconciliation. First, the robust reconciliation model of material balance layer is established to obtain the reconciled values of the first and second feeding quantity and blast velocity. Then, the reconciled values of the material balance layer are used as precise values to construct the robust reconciliation model of heat balance layer. Finally, the reconciled results of the temperature for the blast, gas and calcine can be obtained. Therefore, the hierarchical data reconciliation model based on the novel robust estimation is expressed as follows:

Material balance layer:

$$\min f_1 = \sum_{i=1}^{n} \rho(x_i, \hat{x}_i)$$
$$= \sum_{i=1}^{n} (1 + ((x_i - \hat{x}_i)^2/\sigma_i^2 c_p^{14} - 1)\exp(-((x_i - \hat{x}_i)/\sigma_i c_p)^2))$$
$$= 1 + ((M_1 - \hat{M}_1)^2/\sigma_{M_1}^2 c_p^{14} - 1)\exp(-((M_1 - \hat{M}_1)/\sigma_{M_1} c_p)^2) +$$
$$1 + ((M_2 - \hat{M}_2)^2/\sigma_{M_2}^2 c_p^{14} - 1)\exp(-((M_2 - \hat{M}_2)/\sigma_{M_2} c_p)^2) +$$
$$1 + ((V_0 - \hat{V}_0)^2/\sigma_{V_0}^2 c_p^{14} - 1)\exp(-((V_0 - \hat{V}_0)/\sigma_{V_0} c_p)^2) +$$
$$1 + (\|\rho_1 - \tilde{\rho}_1\|^2/c_p^{14} - 1)\exp(-(\|\rho_1 - \tilde{\rho}_1\|/c_p)^2) \tag{32}$$

$$s.t. \begin{cases} \frac{16}{97}\alpha(\hat{M}_1 + \hat{M}_2) + \hat{V}_0(\rho_0 - \rho_1) = 0 \\ \hat{M}_{1min} \le \hat{M}_1 \le \hat{M}_{1max} \\ \hat{M}_{2min} \le \hat{M}_2 \le \hat{M}_{2max} \\ \hat{V}_{0min} \le \hat{V}_0 \le \hat{V}_{0max} \\ \rho_{1min} \le \rho_1 \le \rho_{1max} \end{cases} \quad (33)$$

where f_1 is the objective function of the robust estimator; x_i and $\hat{x}_i$ represent the measurement and reconciled data for the ith measured variable respectively; $\hat{M}_{1min}$, $\hat{M}_{2min}$, $\hat{V}_{0min}$ and ρ_{1min} denote the bottom bounds of the reconciled results for the first and second feeding quantity, blast velocity and the density of the air, respectively. $\hat{M}_{1max}$, $\hat{M}_{2max}$, $\hat{V}_{0max}$ and ρ_{1max} denote the top bounds of the reconciled results for the first and second feeding quantity, blast velocity and the density of the air, respectively.

Heat balance layer:

$$\begin{aligned} \min f_2 &= \sum_{i=1}^{n} \rho(x_i, \hat{x}_i) \\ &= \sum_{i=1}^{n} (1 + ((x_i - \hat{x}_i)^2/\sigma_i^2 c_p^{14} - 1)\exp(-((x_i - \hat{x}_i)/\sigma_i c_p)^2)) \\ &= 1 + ((T_0 - \hat{T}_0)^2/\sigma_{T_0}^2 c_p^{14} - 1)\exp(-((T_0 - \hat{T}_0)/\sigma_{T_0} c_p)^2) + \\ &\quad 1 + ((T_2 - \hat{T}_2)^2/\sigma_{T_2}^2 c_p^{14} - 1)\exp(-((T_2 - \hat{T}_2)/\sigma_{T2} c_p)^2) + \\ &\quad 1 + ((T_3 - \hat{T}_3)^2/\sigma_{T_3}^2 c_p^{14} - 1)\exp(-((T_3 - \hat{T}_3)/\sigma_{T_3} c_p)^2) + \\ &\quad 1 + (\|Q - \tilde{Q}\|^2/c_p^{14} - 1)\exp(-(\|Q - \tilde{Q}\|/c_p)^2) \end{aligned} \quad (34)$$

$$s.t. \begin{cases} 1000(\hat{M}_1 + \hat{M}_2)T_1 cp_1 + \hat{V}_0\hat{T}_0 cp_2 + \frac{443508}{97}\alpha(\hat{M}_1 + \hat{M}_2)1000 \\ = 1000(1 - \frac{16}{97}\alpha)(\hat{M}_1 + \hat{M}_2)\hat{T}_3 cp_3 + \hat{V}_0\hat{T}_2 cp_0 + Q \\ \hat{T}_{0min} \le \hat{T}_0 \le \hat{T}_{0max} \\ \hat{T}_{2min} \le \hat{T}_2 \le \hat{T}_{2max} \\ \hat{T}_{3min} \le \hat{T}_3 \le \hat{T}_{3max} \\ Q_{min} \le Q \le Q_{max} \end{cases} \quad (35)$$

where f_2 is the objective function of the robust estimator; $\hat{T}_{0min}$, $\hat{T}_{2min}$, $\hat{T}_{3min}$ and Q_{min} denote the bottom bounds of the reconciled results for the temperature of the blast, gas, calcine and heat loss, respectively; $\hat{T}_{0max}$, $\hat{T}_{2max}$, $\hat{T}_{3max}$ and Q_{max} denote the top bounds of the reconciled results for the temperature of the blast, gas, calcine and heat loss, respectively.

Step 4: Reconciled estimation of $\hat{\alpha}$ based on reconciled results and heat balance. The reconciled results obtained in Step 3 are substituted for the objective function of the optimization problem based on heat balance. Then the estimated value of the percentage for ZnS after reconciliation could be obtained. The ZnS estimation model after data reconciliation can be represented as follows:

$$\begin{aligned} \min Y_2 = (&1000(\hat{M}_1 + \hat{M}_2)T_1 cp_1 + \hat{V}_0\hat{T}_0 cp_2 + \frac{443,508}{97}\hat{\alpha}(\hat{M}_1 + \hat{M}_2)1000 \\ &- 1000(1 - \frac{16}{97}\hat{\alpha})(\hat{M}_1 + \hat{M}_2)\hat{T}_3 cp_3 - \hat{V}_0\hat{T}_2 cp_0 - Q)^2 \end{aligned} \quad (36)$$

$$s.t. \hat{\alpha}_{min} \le \hat{\alpha} \le \hat{\alpha}_{max} \quad (37)$$

where Y_2 is the objective of the optimization problem; $\hat{\alpha}$ represents the estimated value of percentage for ZnS after reconciliation; $\hat{\alpha}_{min}$ and $\hat{\alpha}_{max}$ denote the bottom and top bounds of $\hat{\alpha}$ respectively. Through the iterative process of the above steps, greater accuracy and a smaller range of ZnS fluctuation in the feed can be obtained. Therefore, staff can better understand changes of the feeding composition and make adjustments as required.

5. Results from Real Industrial Data

To further demonstrate the effectiveness of the proposed method, 100 steady-state samples are collected from the real industry of a fluidized bed roaster. In addition, the data reconciliation results of the proposed estimator are compared with those of Xie estimator.

The results of the comparison between the two methods are presented in Figures 9–14, which respectively present the reconciled results for the first feeding quantity, the second feeding quantity, blast velocity, blast temperature, gas temperature and calcine temperature. From Figures 9–14, it is known that the reconciled results by the novel robust estimator have a smaller range of fluctuation than that obtained by Xie estimator. It shows that whatever the uncertainty of the fluctuation of the measurements and the number of gross errors they contain, the data reconciliation approach based on the novel robust estimator is highly robust.
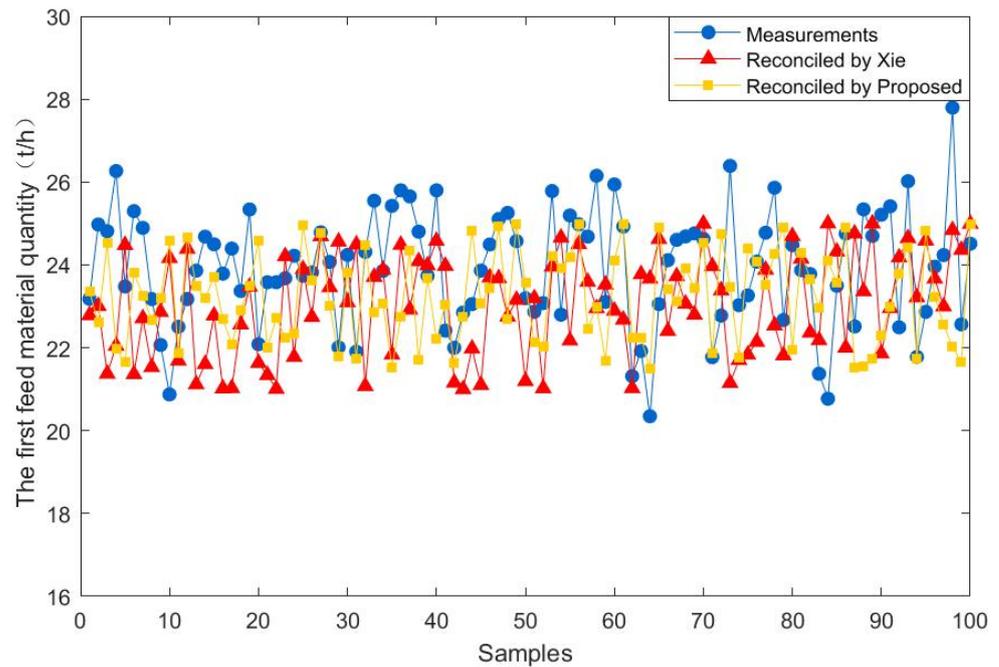
Figure 9. Reconciled results for the first feed material quantity.
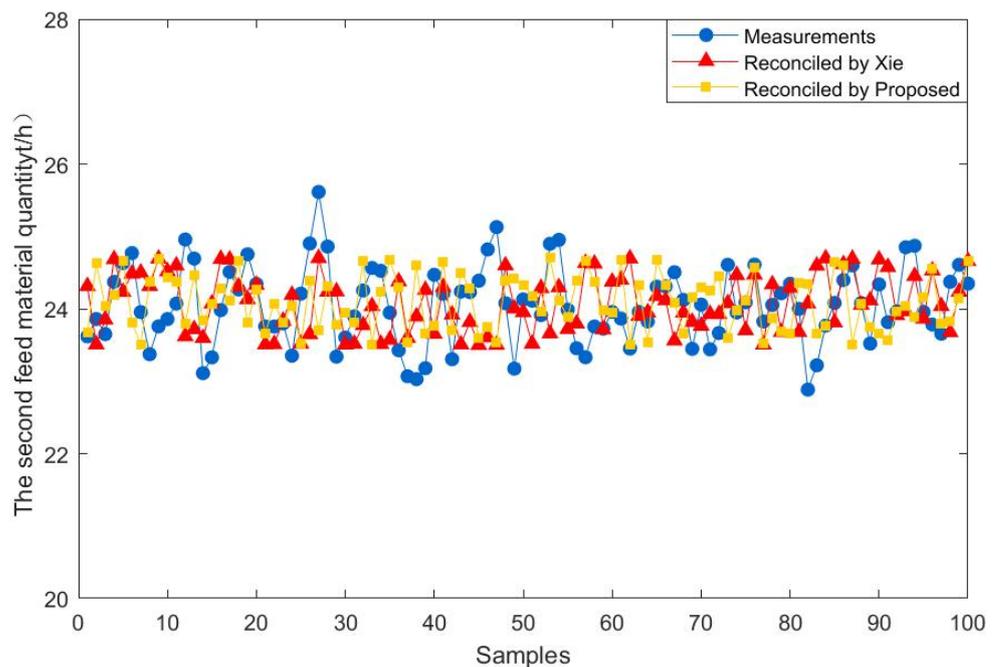
Figure 10. Reconciled results for the second feed material quantity.
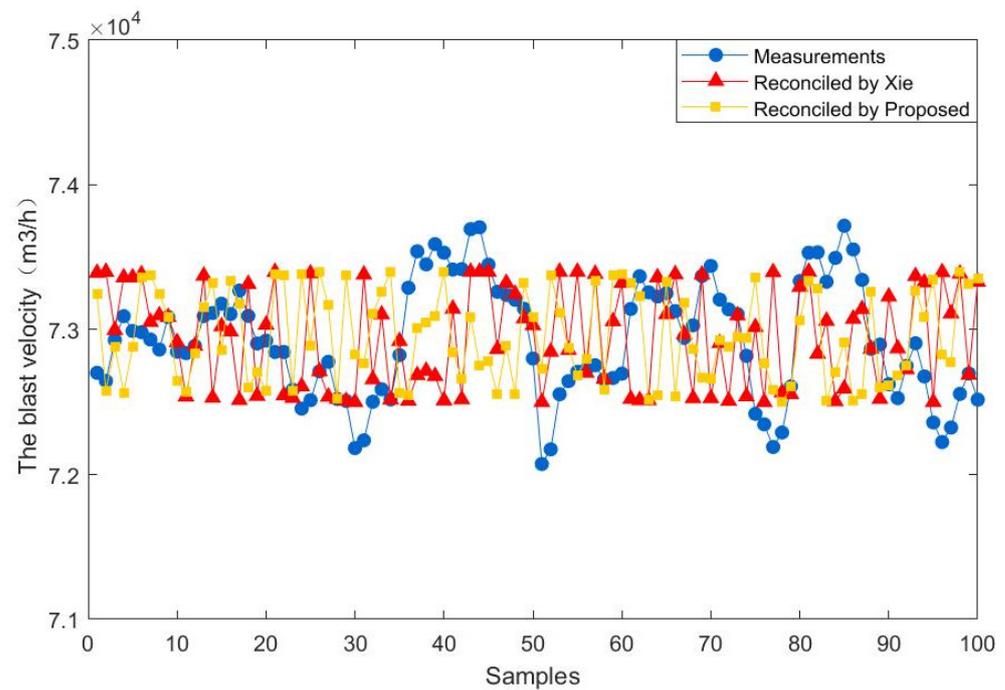
Figure 11. Reconciled results for the blast velocity.
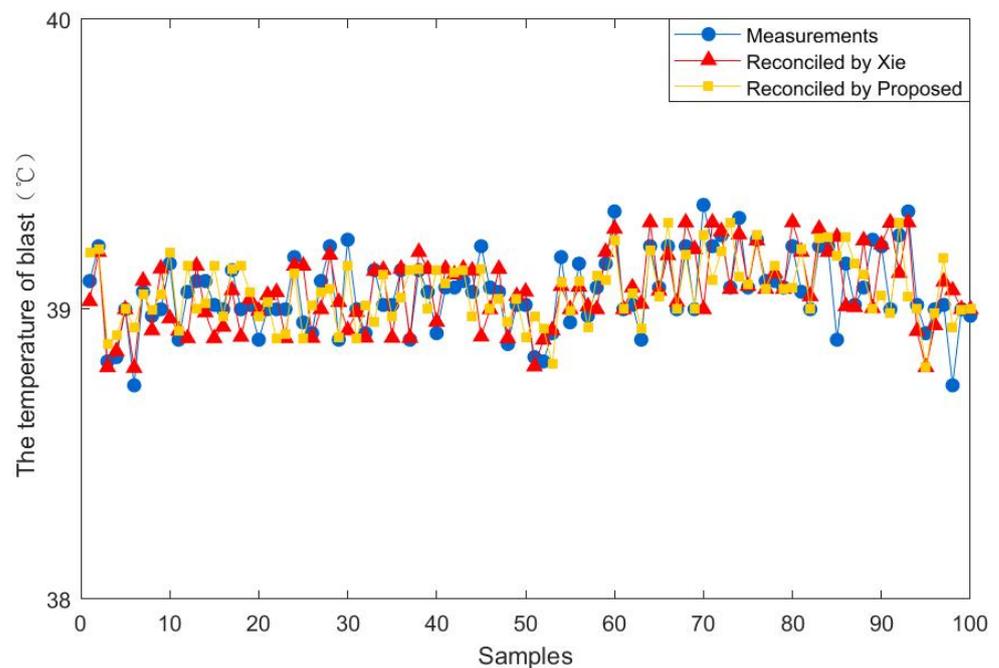
Figure 12. Reconciled results for the temperature of blast.

In addition, the standard deviations of the reconciled results and measurements for six variables obtained from the two methods are calculated, and the results of the comparison are presented in Figure 15. Figure 15 shows that, for the first feeding quantity, the seconding feed quantity, the blast velocity and the temperature of blast, the standard deviation of the reconciled results based on the novel robust estimator is less than Xie. For the temperature of gas and calcine, the standard deviations of reconciled results for the two methods are closer, because the measurement fluctuation ranges for both two variables are all very small. The results demonstrate that the data reconciliation based on the novel robust estimator provides better reconciliation results for variables with a large fluctuation range.
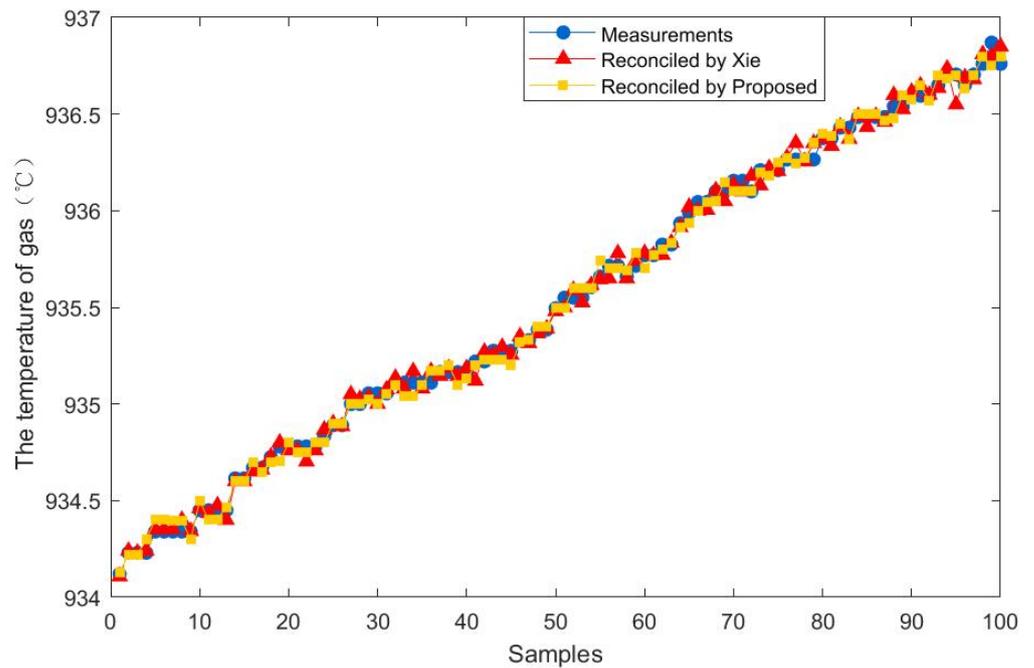
Figure 13. Reconciled results for the temperature of gas.
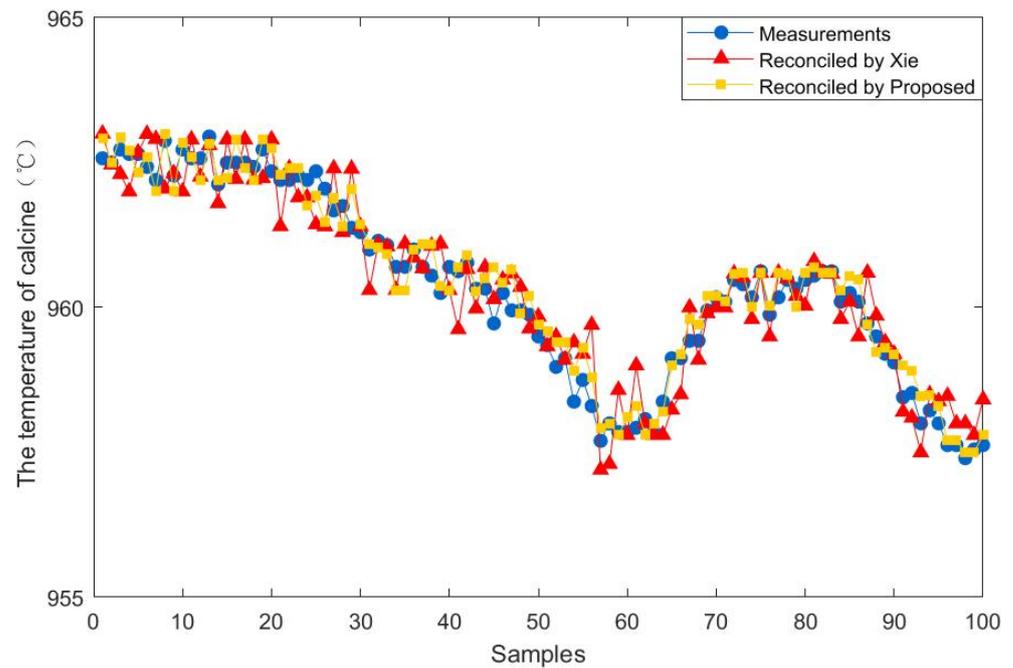

Figure 14. Reconciled results for the temperature of calcine.

For the change in feeding composition, the above 100 steady-state samples are also collected to compare the percentage of ZnS before and after data reconciliation. Experimental results are shown in Figures 16 and 17. It is evident that the ZnS fluctuation range obtained by the proposed method is smaller than that before data reconciliation. It illustrates that the percentage of ZnS estimated by the reconciled results is more accurate and could better reflect the change of feeding composition over a period of time.
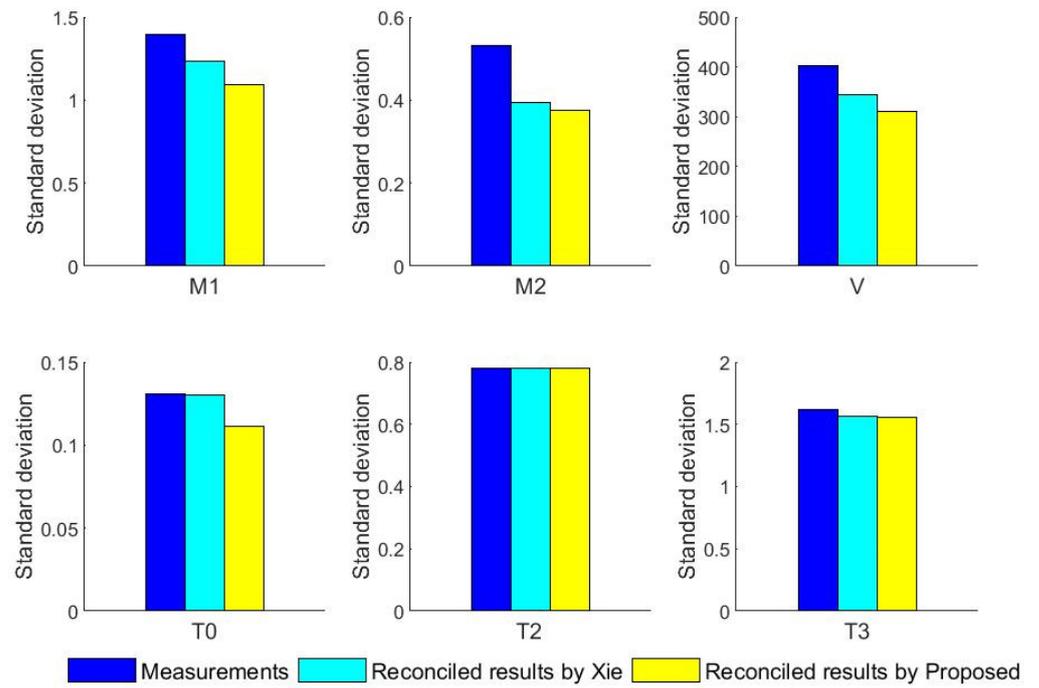
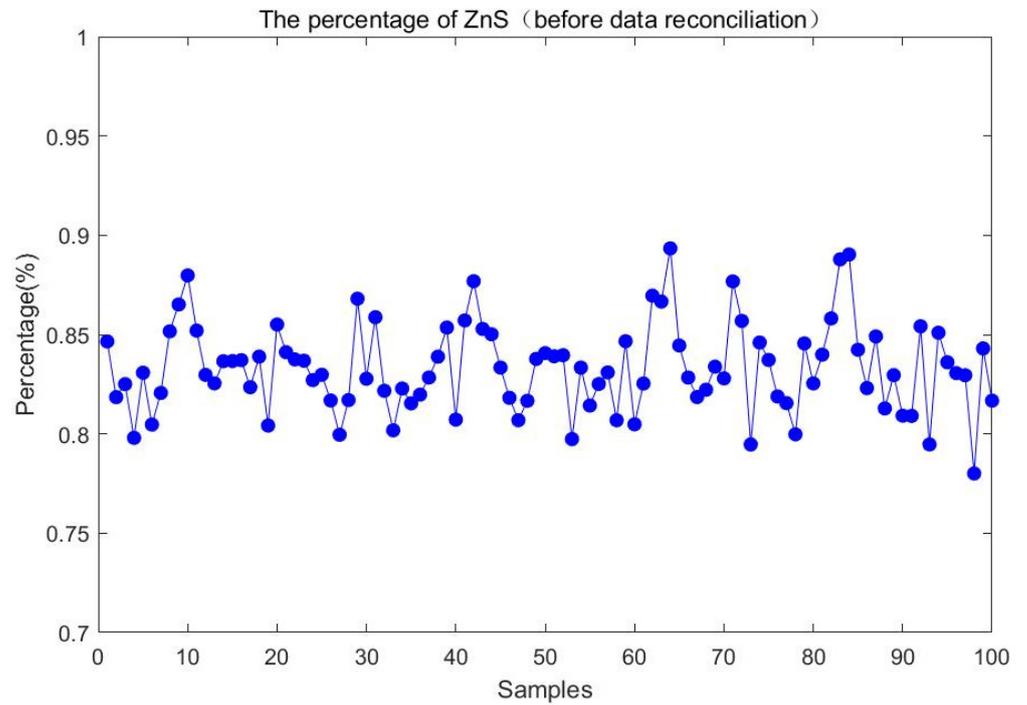**Figure 15.** The standard deviation of reconciled results and measurements.

Figure 16. The percentage of ZnS before data reconciliation.
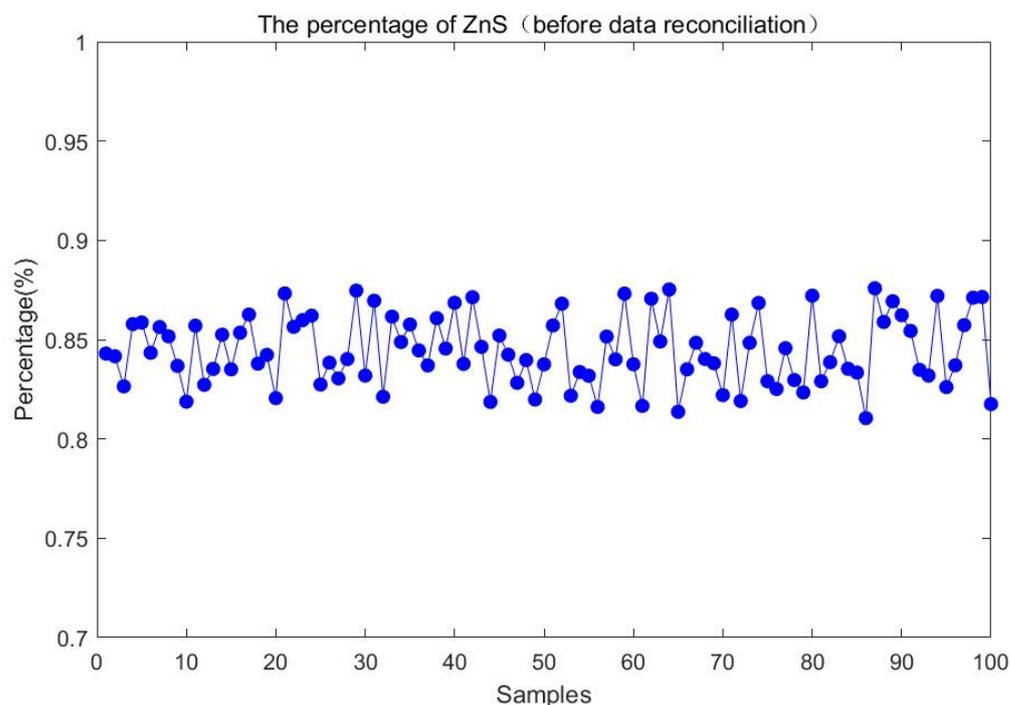
Figure 17. The percentage of ZnS after data reconciliation.

6. Conclusions

The feeding composition in an industrial process can better reflect the state of production. In this paper, estimation of feeding composition based on robust data reconciliation is proposed to account for the unknown change in feeding composition. According to the robust estimation theory, a novel robust estimator is presented to handle the measurements with random errors and gross errors. In comparison with the objective function and influence function of other robust estimators, the proposed estimator is proved to be more robust. Besides, the linear and nonlinear numerical examples are used to further verify the effectiveness of the estimator. Then, an iterative robust hierarchical data reconciliation and estimation of feeding composition strategy is proposed and used for a fluidized bed roaster. The reconciled and estimated results show that the proposed strategy is beneficial for the estimation of feeding composition in the actual industrial process.

Data Availability Statement: Publicly available datasets were analyzed in this study. It can be found here: https://open.library.ubc.ca/cIRcle/collections/ubctheses/831/items/1.0078522 (accessed on 1 February 2021).

References

1. Kuehn, D.R.; Davidson, H. Computer control II. Mathematics of control. *Chem. Eng. Prog.* **1961**, *57*, 44–47.
2. Narasimhan, S.; Jordache, C. *Data Reconciliation and Gross Error Detection: An Intelligent Use of Process Data*; Elsevier: Amsterdam, The Netherlands, 1999.

3. Prata, D.M.; Lima, E.L.; Pinto, J.C. In-Line Monitoring of Bulk Polypropylene Reactors Based on Data Reconciliation Procedures. *Macromol. Symp.* **2008**, *271*, 26–37. [CrossRef]

4. Prata, D.M.; Lima, E.L.; Pinto, J.C. Nonlinear Dynamic Data Reconciliation in Real Time in Actual Processes. *Comput. Aided Chem. Eng.* **2009**, *27*, 47–54.

5. Abu-El-Zeet, Z.H.; Roberts, P.D.; Becerra, V.M. Enhancing model predictive control using dynamic data reconciliation. *Alche J.* **2002**, *48*, 324–333. [CrossRef]

6. Sun, B.; Yang, C.; Wang, Y.; Gui, W.; Craig, I.; Olivier, L. A comprehensive hybrid first principles/machine learning modeling framework for complex industrial processes. *J. Process Control* **2020**, *86*, 30–43. [CrossRef]

7. Faber, R.; Arellano-Garcia, H.; Li, P.; Wozny, G. An optimization framework for parameter estimation of large-scale systems. *Chem. Eng. Process.* **2007**, *46*, 1085–1095. [CrossRef]

8. Schladt, M.; Hu, B. Soft Sensor Based on Nonlinear Steady-State Data Reconciliation in the Process Industry. *Chem. Eng. Process.* **2007**, *46*, 1107–1115. [CrossRef]

9. Su, Q.; Bommireddy, Y.; Shah, Y.; Ganesh, S.; Moreno, M.; Liu, J.; Gonzalez, M.; Yasdanpanah, N.; O'Connor, T.; Reklaitis, G.V.; et al. Data reconciliation in the Quality-byDesign (QbD) implementation of pharmaceutical continuous tablet manufacturing. *Int. J. Pharm.* **2019**, *563*, 259–272. [CrossRef]

10. Chiari, M.; Bussani, G.; Grotolli, M.G.; Pierucci, S. On-line Data Reconciliation and Optimisation: Refinery Applications. *Comput. Chem. Eng.* **1997**, *21*, s1185–s1190. [CrossRef]

11. Lee, M.H.; Lee, S.J.; Han, C.; Chang, K.S.; Kim, S.H.; Sang, H.Y. Hierarchical on-line data reconciliation and optimization for an industrial utility plant. *Comput. Chem. Eng.* **1998**, *22*, s247–s254. [CrossRef]

12. Almásy, G.A.; Sztanó, T. Checking and correction of measurements on the basis of linear system model. *Probl. Control Inf. Theory* **1975**, *4*, 57–69.

13. Mah, R.S.H.; Tamhane, A.C. Detection of gross errors in process data. *Aiche J.* **2010**, *28*, 828–830. [CrossRef]

14. Mah, R.S.; Stanley, G.M.; Downing, D.M. Reconciliation and rectification of process flow and inventory data. *Ind. Eng. Chem. Process Des. Dev.* **1976**, *15*, 175–183. [CrossRef]

15. Narasimhan, S.; Mah, R.S.H. Generalized likelihood ratio method for gross error identification. *Aiche J.* **1987**, *33*, 1514–1521. [CrossRef]

16. Tong, H.; Crowe, C.M. Detection of gross erros in data reconciliation by principal component analysis. *Aiche J.* **2010**, *41*, 1712–1722. [CrossRef]

17. Yu, M.; Hong-Ye, S.U.; Jian, C.H.U. A support vector regression approach for recursive simultaneous data reconciliation and gross error detection in nonlinear dynamical systems. *Acta Autom. Sin.* **2009**, *35*, 707–716.

18. Zhang, Z.; Chen, J. Simultaneous data reconciliation and gross error detection for dynamic systems using particle filter and measurement test. *Comput. Chem. Eng.* **2014**, *69*, 66–74. [CrossRef]

19. Yuan, Y.; Khatibisepehr, S.; Huang, B.; Li, Z. Bayesian method for simultaneous gross error detection and data reconciliation. *Aiche J.* **2015**, *61*, 3232–3248. [CrossRef]

20. Tjoa, I.B.; Biegler, L.T. Simultaneous strategies for data reconciliation and gross error detection of nonlinear systems. *Comput. Chem. Eng.* **1991**, *15*, 679–690. [CrossRef]

21. Johnson, L.P.M.; Kramer, M.A. Maximum likelihood data rectification: Steady-state systems. *Alche J.* **1995**, *41*, 2415–2426. [CrossRef]

22. Arora, N.; Biegler, L.T. Redescending estimators for data reconciliation and parameter estimation. *Comput. Chem. Eng.* **2001**, *25*, 1585–1599. [CrossRef]

23. Wang, D.; Romagnoli, J.A. A Framework for Robust Data Reconciliation Based on a Generalized Objective Function. *Ind. Eng. Chem. Res.* **2003**, *42*, 3075–3084. [CrossRef]

24. Özyurt, D.B.; Pike, R.W. Theory and practice of simultaneous data reconciliation and gross error detection for chemical processes. *Comput. Chem. Eng.* **2004**, *28*, 381–402. [CrossRef]

25. Ragot, J.; Chadli, M.; Maquin, D. Mass balance equilibration: A robust approach using contaminated distribution. *Aiche J.* **2005**, *51*, 1569–1575. [CrossRef]

26. Prata, D.M.; Schwaab, M.; Lima, E.L.; Pinto, J.C. Simultaneous robust data reconciliation and gross error detection through particle swarm optimization for an industrial polypropylene reactor. *Chem. Eng. Sci.* **2010**, *65*, 4943–4954. [CrossRef]

27. Zhang, Z.; Shao, Z.; Chen, X.; Wang, K.; Qian, J. Quasi-weighted least squares estimator for data reconciliation. *Comput. Chem. Eng.* **2010**, *34*, 154–162. [CrossRef]

28. Llanos, C.E.; Sanchéz, M.C.; Maronna, R.A. Robust estimators for data reconciliation. *Ind. Eng. Chem. Res.* **2015**, *54*, 5096–5105. [CrossRef]

29. Alighardashi, H.; Magbool Jan, N.; Huang, B. Expectation Maximization Approach for Simultaneous Gross Error Detection and Data Reconciliation Using Gaussian Mixture Distribution. *Ind. Eng. Chem. Res.* **2017**, *56*, 14530–14544. [CrossRef]

30. Xie, S.; Yang, C.; Yuan, X.; Wang, X.; Xie, Y. A novel robust data reconciliation method for industrial processes. *Control Eng. Pract.* **2019**, *83*, 203–212. [CrossRef]

31. de Menezes, D.Q.F.; Prata, D.M.; Secchi, A.R.; Pinto, J.C. A review on robust M-estimators for regression analysis. *Comput. Chem. Eng.* **2021**, *147*, 107254. [CrossRef]

32. Rey, W.J.J. *Introduction to Robust and Quasi-Robust Statistical Methods*; Springer Science & Business Media: Berlin, Germany, 2012.

33. Fair, R.C. On the robust estimation of econometric models. *Ann. Econ. Soc. Meas.* **1974**, *3*, 667.
34. Jin, S.Y.; Li, X.W.; Huang, Z.J.; Meng, L. A new target function for robust data reconciliation. *Ind. Eng. Chem. Res.* **2012**, *51*, 10220–10224. [CrossRef]
35. Hoaglin, D.C.; Mosteller, F.; Tukey, J.W. *Understanding Robust and Exploratory Data Analysis*; Wiley: New York, NY, USA, 1983; Volume 3.
36. Albuquerque, J.S.; Biegler, L.T. Data reconciliation and gross-error detection for dynamic systems. *Aiche J.* **1996**, *42*, 2841–2856. [CrossRef]
37. Shevlyakov, G.; Morgenthaler, S.; Shurygin, A. Redescending m-estimators. *J. Stat. Plann. Inference* **2008**, *138*, 2906–2917. [CrossRef]
38. Zhou, X.J.; Yang, C.H.; Gui, W.H. State transition algorithm. *J. Ind. Manag. Optim.* **2012**, *8*, 1039–1056. [CrossRef]