Bayesian Method for Simultaneous Gross Error Detection and Data Reconciliation

Yuan Yuan, Shima Khatibisepehr, Biao Huang, and Zukui Li
Dept. of Chemical and Materials Engineering, University of Alberta, Edmonton, AB, Canada, T6G2G6

DOI 10.1002/aic.14864
Published online May 23, 2015 in Wiley Online Library (wileyonlinelibrary.com)

Process measurements collected from daily industrial plant operations are essential for process monitoring, control, and optimization. However, those measurements are generally corrupted by errors, which include gross errors and random errors. Conventionally, those two types of errors were addressed separately by gross error detection and data reconciliation. Solving the simultaneous gross error detection and data reconciliation problem using the hierarchical Bayesian inference technique is focused. The proposed approach solves the following problems in a unified framework. First, it detects which measurements contain gross errors. Second, the magnitudes of the gross errors are estimated. Third, the covariance matrix of the random errors is estimated. Finally, data reconciliation is performed using the maximum a posteriori estimation. The proposed algorithm is applicable to both linear and nonlinear systems. For nonlinear case, the algorithm does not involve any linearization or approximation steps. Numerical case studies are provided to demonstrate the effectiveness of the proposed method. © 2015 American Institute of Chemical Engineers *AIChE J*, 61: 3232–3248, 2015

Keywords: gross error detection, data reconciliation, Bayesian inference, serial strategy

Introduction

Process data measurements are of great importance for model identification, process control, online optimization, and process monitoring. However, process measurements are inevitably contaminated by errors. There are two types of errors that contribute to the total error in a measurement: random errors can be referred to as noise and they exist in any measurements; gross errors are nonrandom errors caused by malfunctioning, miscalibration, fault of sensors, and so forth. To obtain reliable process data, it is necessary to remove the errors from the measurements. Usually, random errors and gross errors are addressed separately by data reconciliation and gross error detection.

Data reconciliation is used to enhance the accuracy of measurements by reducing the influence of random errors. Its objective is to estimate the true value of the measurements that satisfy the process models. In general, data reconciliation is based on the assumption that there are only random errors in the measurements. The data reconciliation problem can be formulated as an optimization problem.[1] Mah[2] provided an overview of the basis of data reconciliation.

The existence of gross errors will invalidate the statistical basis of data reconciliation. Hence, gross errors must be detected and removed/compensated before data reconciliation. Several statistical tests have been utilized for gross error detection in linear model under steady state, such as global test,[3] measurement test,[4,5] nodal test,[3,6] generalized likelihood ratios,[7] Bonferroni tests,[8] principle component tests,[9] and so forth. In practice, we are not only interested in detecting the presence of gross errors but also want to identify the locations of gross errors as well as to estimate the sizes of them. These targets can be fulfilled by different strategies combined with statistical tests. Serial elimination strategy[10–12] detects each measurement one by one, recomputes the statistical tests and then finally eliminates the gross error candidates (GEC). Serial compensation strategy[7] can be used to estimate the sizes of gross errors. In this strategy, gross errors are estimated and the measurements are compensated in turn. It is applicable to all kinds of gross errors but its results rely on the accuracy of estimated size of gross errors.[8] Collective compensation strategy[8,13,14] has also been proposed to estimate all gross errors simultaneously. This method is more accurate than others.[14,15] However, it is computationally expensive.

There is also a close relationship between the problem of data reconciliation and gross error detection. The residuals obtained from a data reconciliation step are generally used in the statistical test-based gross error detection. Some gross error detection techniques do not depend on the residuals, however, measurements containing gross errors have to be eliminated or compensated before data reconciliation. As gross error detection and data reconciliation are closely related, many researchers have proposed methods to combine the gross error detection and data reconciliation and address them simultaneously.[16,17] The objective is to remove the random errors as well as the gross errors and to obtain clean data.

Although component mass balance or energy balance is included in the process model, nonlinear data reconciliation problem need to be addressed. The usual procedure is to first perform a linearization of the process model. Sequential

Correspondence concerning this article should be addressed to B. Huang at biao.huang@ualberta.ca.

quadratic programming technique solves a nonlinear optimization problem by successively solving a series of quadratic programming problems.[18–20] Tjoa and Biegler used a new distribution function considering both contributions form random errors and gross errors and proposed a hybrid sequential quadratic programming method to solve the nonlinear gross error detection and data reconciliation problem simultaneously.[16]

This article is devoted to the issue of combining gross error detection and identification problem with data reconciliation problem within a hierarchical Bayesian framework. Hierarchical Bayesian framework has been applied to various problems. For instance, MacKay[21] first proposed the heuristic Bayesian evidence framework and this framework was applied to neural network modeling by MacKay.[22] The hierarchical Bayesian procedure was used to address the image modeling and restoration problem by Molina et al.[23] and Galatsanos et al.[24] Kwok[25] and Suykens et al.[26] derived the probabilistic formulation of the least squares support vector machine within a hierarchical Bayesian evidence framework. The hierarchical Bayesian framework was utilized for process identification with outliers in the dataset by Khatibisepehr and Huang.[27]

This article focuses on the system working at steady state which can be both linear and nonlinear. New strategies based on hierarchical Bayesian combining with serial identification and collective estimation of gross errors are proposed. Instead of using statistical tests method, (e.g., Sánchez et al.[14] and Jiang and Bagajewicz[28]), the proposed approach compares the conditional probability ratio of containing gross error and not containing gross error, so as to identify a list of suspect GEC. The proposed algorithm performs gross error detection and data reconciliation simultaneously and also estimates the magnitudes of the gross errors. Furthermore, the proposed approach does not assume that the covariance matrix of random errors is known. Instead, the proposed approach estimates the covariance matrix.

The rest of the article is organized as follows. In the next section, we first provide the problem statement. Then, the simultaneous gross error detection and data reconciliation problem is formulated in a hierarchical Bayesian framework, where different variables are estimated in different layers. The gross error exact detectability issue will be discussed and a serial strategy is adopted for gross error identification. Next, numerical examples are provided to demonstrate the effectiveness of the proposed approach for both linear and nonlinear cases, followed by conclusion.

Problem Statement

The objective of simultaneous gross error detection and data reconciliation can be stated as follows. First, detect which measurements contain gross errors; second, estimate the magnitudes of gross errors as well as the covariance matrix of the random errors; finally, apply data reconciliation to estimate the correct value of the data.

In this article, the system under consideration is assumed to be time invariant and operating under steady state. According to mass balance or energy balance, the system model can be expressed as

$$f(x,u)=0$$

where x is an $n\times 1$ vector of true values of measured variables, n is the number of measured variables, and u is a vector of unmeasured variables.

In the real process, the process data are automatically sampled and recorded at regular time intervals. It is assumed that a gross error persists in a measurement (if it exists) during a measurement period consisting of m sampling points. According to the above assumptions, the measurement data are organized as follows

$$D=\begin{pmatrix} y_{11} & y_{12} & \dots & y_{1m} \\ y_{21} & y_{22} & \dots & y_{2m} \\ \dots & \dots & \dots & \dots \\ y_{n1} & y_{n2} & \dots & y_{nm} \end{pmatrix}=\begin{pmatrix} Y_1{}^C & Y_2{}^C & \dots & Y_m{}^C \end{pmatrix}=\begin{pmatrix} Y_1^R \\ Y_2^R \\ \dots \\ Y_n^R \end{pmatrix}$$

where Y_j^C, $j=1,\dots,m$, is the data at each sampling point and Y_i^R, $i=1,\dots,n$, is the data for each variable. The data are assumed to be mutually independent.

At each sampling point, the measurement model can be described as

$$Y_j^C=x+\eta\delta+\varepsilon,\quad j=1,\dots,m$$

where η is an $n\times n$ matrix which indicates whether there is a gross error in the measurement or not (0 means no gross error and 1 means gross error), δ is an $n\times 1$ vector which denotes the magnitudes of the biases, ε is the random error vector which follows a multivariate normal distribution, that is, $\varepsilon\sim N(0,\Sigma)$, and Σ is the $n\times n$ covariance matrix. Notice that η is constructed as a diagonal matrix with the elements on the diagonal to be the gross error indicators for each measurements

$$\eta=\begin{pmatrix} \eta_1 & 0 & \dots & 0 \\ 0 & \eta_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \eta_n \end{pmatrix}$$

As the measurements are obtained from different instruments, we can assume that the measurement noise is mutually independent. Then, Σ is a diagonal matrix

$$\Sigma=\begin{pmatrix} \sigma_{11}^2 & 0 & \dots & 0 \\ 0 & \sigma_{22}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{nn}^2 \end{pmatrix}$$

For the purpose of simplification, the precision is introduced as $\alpha_i=\sigma_{ii}^{-2}$.

It is known that finding estimates of parameters using maximum likelihood estimation may lead to an ill-posed problem as the estimation can be underdetermined and is also sensitive to the noise in the data.[21] A useful way to avoid that is to combine the information from the data with some additional knowledge concerning the distribution of the parameters called prior distribution. In this article, we set priors for the magnitudes of biases δ and the correct values of measurements x. Specifically, a uniform distribution is introduced as the prior distribution of δ. A normal distribution is introduced as the prior distribution of x, that is, $x\sim N(\mu_0,\Sigma_0)$, where hyperparameter μ_0 is a $n\times 1$ vector and Σ_0 is the covariance matrix. If any hyperparameter of those prior distributions is not known *a priori*, its value can be estimated by an intermediate step of the whole process. For instance, the hyperparameters μ_0 and Σ_0 are not known *a priori* in this article. The estimation of x in

the current iteration is taken as the estimation of μ_0 in the next step and Σ_0 is estimated by maximizing the posterior distribution over it. In this work, we assume the correct values of x are independent and Σ_0 is a diagonal matrix

$$\Sigma_0 = \begin{pmatrix} \sigma_{11,0}^2 & 0 & \dots & 0 \\ 0 & \sigma_{22,0}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{nn,0}^2 \end{pmatrix}$$

and the precision for $\sigma_{ii,0}$ is introduced as $\alpha_{i0} = \sigma_{ii,0}^{-2}$.

Proposed Hierarchical Bayesian Framework

In the simultaneous gross error detection and data reconciliation problem, five different parameters (x, δ, η, Σ, Σ_0) are considered. In the proposed Bayesian framework, maximum *a posteriori* (MAP) is used to estimate the parameters. To obtain MAP estimates for the five parameters simultaneously, the joint probability density function $P(x, \delta, \Sigma, \Sigma_0, \eta|D)$ should be maximized. However, maximizing such a posterior probability density functions is complex. The corresponding optimization problem is difficult to solve directly. To avoid the difficulties in direct maximization of $P(x, \delta, \Sigma, \Sigma_0, \eta|D)$, a layered solution framework is proposed. Based on the chain rule, the joint probability density function is factorized into three parts

$$P(x, \delta, \Sigma, \Sigma_0, \eta|D) = P(x, \delta|\Sigma, \Sigma_0, \eta, D) P(\Sigma, \Sigma_0|\eta, D) P(\eta|D)$$

The proposed method includes three levels according to the above factorization. The correct values of the measurements x and the magnitudes of gross errors δ are estimated at Level 1. The covariance matrix Σ of the random error and the hyperparamter Σ_0 are estimated at Level 2 and the gross error indicator matrix η is addressed at Level 3. The overall procedure of the Bayesian algorithm is summarized as follows

Level 1: Inference of x and δ

$$\max_{x,\delta} P(x, \delta|\Sigma, \Sigma_0, \eta, D) = \max_{x,\delta} \frac{P(D|x, \delta, \Sigma, \Sigma_0, \eta) P(x, \delta|\Sigma, \Sigma_0, \eta)}{P(D|\Sigma, \Sigma_0, \eta)}$$

Level 2: Inference of Σ and Σ_0

$$\max_{\Sigma,\Sigma_0} P(\Sigma, \Sigma_0|\eta, D) = \max_{\Sigma,\Sigma_0} \frac{P(D|\Sigma, \Sigma_0, \eta) P(\Sigma, \Sigma_0|\eta)}{P(D|\eta)}$$

Level 3: Inference of η

$$\max_{\eta} P(\eta_i|D) = \max_{\eta} \frac{P(D|\eta) P(\eta)}{P(D)} = \max_{\eta} P(D|\eta) P(\eta)$$

Each level of the above Bayesian framework has the following form

$$\text{Posterior} = \frac{\text{Likelihood}}{\text{Evidence}} \times \text{Prior}$$

It is easy to see that the likelihood at a certain level is equal to the evidence at the previous level. In this way, the three levels are connected to each other. The procedure is iterated until convergence.

First layer: Inference of x and δ

Given the collected data D, the hyperparameters Σ and Σ_0, and the matrix of gross error indicators η, the MAP estimates of the correct values of measurements x, and the magnitudes of the biases δ are obtained by maximizing the posterior density function $P(x, \delta|\Sigma, \Sigma_0, \eta, D)$. Using Bayes rule, the formulation in the first layer of the hierarchical framework is

$$P(x, \delta|\Sigma, \Sigma_0, \eta, D) = \frac{P(D|x, \delta, \Sigma, \Sigma_0, \eta) P(x, \delta|\Sigma, \Sigma_0, \eta)}{P(D|\Sigma, \Sigma_0, \eta)} \quad (1)$$

As stated in the above section, the following priori distributions for x and δ are slected: a normal distribution is introduced for x and a uniform distribution is set for δ. It is also assumed that x and δ are independent. For the correct value of x, it can be considered to be independent of Σ and η, and dependent on its hyperparameter Σ_0. The magnitudes δ of the gross errors can be considered to be independent of Σ, η, and Σ_0

$$\begin{aligned} P(x, \delta|\Sigma, \Sigma_0, \eta) &= P(x|\Sigma_0) P(\delta) \\ &= (2\pi)^{-n/2} |\Sigma_0|^{-1/2} \exp\left\{ -\frac{1}{2} (x-\mu_0)^T \Sigma_0^{-1} (x-\mu_0) \right\} \end{aligned} \quad (2)$$

Given x and δ, the sampled data D would be independent of hyperparameter Σ_0, and the likelihood can be directly obtained from the measurement model, that is, $P(D|x, \delta, \Sigma, \Sigma_0, \eta) = P(D|x, \delta, \Sigma, \eta)$

$$\begin{aligned} P(D|x, \delta, \Sigma, \Sigma_0, \eta) = P(D|x, \delta, \Sigma, \eta) &= \prod_{j=1}^{m} P(Y_j^C|x, \delta, \Sigma, \eta) \\ &= (2\pi)^{-mn/2} |\Sigma|^{-m/2} \\ &\times \exp\left\{ -\frac{1}{2} \sum_{j=1}^{m} (Y_j^C - x - \eta\delta)^T \Sigma^{-1} (Y_j^C - x - \eta\delta) \right\} \end{aligned} \quad (3)$$

The posterior probability of x and δ can be calculated by combining Eqs. 2 and 3

$$\begin{aligned} &P(x, \delta|\Sigma, \Sigma_0, \eta, D) \propto P(D|x, \delta, \Sigma, \Sigma_0, \eta) P(x, \delta|\Sigma, \Sigma_0, \eta) \\ &= (2\pi)^{-(m+1)n/2} |\Sigma|^{-m/2} |\Sigma_0|^{-1/2} \\ &\times \exp\left\{ -\frac{1}{2} \left[\sum_{j=1}^{m} (Y_j^C - x - \eta\delta)^T \Sigma^{-1} (Y_j^C - x - \eta\delta) \right.\right. \\ &\left.\left. + (x-\mu_0)^T \Sigma_0^{-1} (x-\mu_0) \right] \right\} \propto \exp\{-J_1(x, \delta)\} \end{aligned} \quad (4)$$

where

$$\begin{aligned} J_1(x, \delta) = \frac{1}{2} \left[\sum_{j=1}^{m} (Y_j^C - x - \eta\delta)^T \Sigma^{-1} (Y_j^C - x - \eta\delta) \right. \\ \left. + (x-\mu_0)^T \Sigma_0^{-1} (x-\mu_0) \right] \end{aligned} \quad (5)$$

All positive constants in Eq. 4 are neglected, as they do not affect the optimal solution of MAP problem. To estimate the most probable values of x and δ, x^{MP} and δ^{MP}, the posterior probability should be maximized, or equivalently, the negative logarithm of Eq. 4 should be minimized

$$\max_{x,\delta} P(x, \delta|\Sigma, \Sigma_0, \eta, D) = \min_{x,\delta} J_1(x, \delta) \quad (6)$$

As the correct value of the measurement x must satisfy the material balance, constraints $f(x, u) = 0$ are added to this optimization problem. Then, the optimization problem for estimating x and δ becomes

$$\begin{aligned} &\min_{x,\delta} J_1(x, \delta) \\ &\text{s.t. } f(x, u) = 0 \end{aligned} \quad (7)$$

Linear Case. Assume the system model with only measured variables is represented as $A_x x = 0$ (matrix A_x is obtained

by decomposing the measured and unmeasured parts). Lagrange multipliers can be introduced to the above optimization problem. The observable unmeasured variables can be calculated through the system model after the gross error detection and the redundant measured variables are reconciled.

$$L=J_1(x,\delta)+\lambda^T A_x x \tag{8}$$

Based on the following optimality condition

$$\frac{\partial L}{\partial x}=(x-\mu_0)^T{\Sigma_0}^{-1}-\sum_{j=1}^{m}(Y_j-x-\eta\delta)^T\Sigma^{-1}+\lambda^T A_x=0 \tag{9}$$

$$\frac{\partial L}{\partial \delta}=-\sum_{j=1}^{m}(Y_j-x-\eta\delta)^T\Sigma^{-1}=0 \tag{10}$$

$$\frac{\partial L}{\partial \lambda}=A_x x=0 \tag{11}$$

The analytical expression for the estimations, x^{MP} and δ^{MP}, can be derived

$$x^{\mathrm{MP}}=(I-mR(I-W)\Sigma^{-1}\eta)^{-1}R(I-W)\left({\Sigma_0}^{-1}\mu_0+\Sigma^{-1}\sum_{j=1}^{m}Y_j^C-\Sigma^{-1}\eta\sum_{j=1}^{m}Y_j^C\right) \tag{12}$$

$$\delta^{\mathrm{MP}}=\eta\left(\frac{1}{m}\sum_{j=1}^{m}Y_j^C-x\right) \tag{13}$$

where $R=({\Sigma_0}^{-1}+m\Sigma^{-1})^{-1}$ and $W=A_x^T[A_xRA_x^T]^{-1}A_xR$.

Nonlinear Case. If component balance and energy balance are considered, nonliear model can be obtained. If analytical solutions for the optimization problem in Eq. 7 cannot be derived, numerical methods can be used to solve the nonlinear optimization problem to get solution for x^{MP} and δ^{MP}. The unmeasured variables can be calculated through the system model $f(x,u)=0$ with the values of x^{MP}.

Second layer: Inference of hyperparameters Σ and Σ_0

Hyperparameters Σ and Σ_0 can be estimated in the second layer. The posterior distribution of the hyperparameters is expressed as

$$P(\Sigma,\Sigma_0|\eta,D)=\frac{P(D|\Sigma,\Sigma_0,\eta)P(\Sigma,\Sigma_0|\eta)}{P(D|\eta)} \tag{14}$$

Based on the assumption that the measurements are independent of each other, the likelihood can be separated and the posterior distribution corresponding to each measurement can be written as

$$P(\sigma_{ii}^2,\sigma_{ii,0}^2|\eta_i,Y_i^R)=\frac{P(Y_i^R|\sigma_{ii}^2,\sigma_{ii,0}^2,\eta_i)P(\sigma_{ii}^2,\sigma_{ii,0}^2|\eta_i)}{P(Y_i^R|\eta_i)} \tag{15}$$

For prior distributions, it is assumed that σ_{ii}^2 and $\sigma_{ii,0}^2$ are independent, so $P(\sigma_{ii}^2,\sigma_{ii,0}^2|\eta_i)=P(\sigma_{ii}^2|\eta_i)\times P(\sigma_{ii,0}^2|\eta_i)$. Using the precisions defined for σ_{ii}^2 and $\sigma_{ii,0}^2$, we have $P(\sigma_{ii}^2,\sigma_{ii,0}^2|\eta_i)=P(\alpha_i,\alpha_{i0}|\eta_i)=P(\alpha_i|\eta_i)P(\alpha_{i0}|\eta_i)$. If there is no *a priori* information for the hyperparameters, a uniform distribution can be used to describe appropriate noninformative priors. To incorporate the prior knowledge, conjugate priors are commonly assigned so that the resulting posterior distribution can be conveniently evaluated and an analytical solution can be obtained. In this layer, gamma distributions are considered as hyperpriors

$$P(\alpha_i|\eta_i)=\frac{{s_i}^{k_i}{\alpha_i}^{k_i-1}}{\Gamma(k_i)}\exp(-s_i\alpha_i)\propto {\alpha_i}^{k_i-1}\exp(-s_i\alpha_i) \tag{16}$$

$$P(\alpha_{i0}|\eta_i)=\frac{{s_{i0}}^{k_{i0}}{\alpha_{i0}}^{k_{i0}-1}}{\Gamma(k_{i0})}\exp(-s_{i0}\alpha_{i0})\propto {\alpha_{i0}}^{k_{i0}-1}\exp(-s_{i0}\alpha_{i0}) \tag{17}$$

$$P(\alpha_i,\alpha_{i0}|\eta_i)=P(\alpha_i|\eta_i)P(\alpha_{i0}|\eta_i)\propto {\alpha_i}^{k_i-1}\exp(-s_i\alpha_i){\alpha_{i0}}^{k_{i0}-1}\exp(-s_{i0}\alpha_{i0}) \tag{18}$$

where k_j is the shape parameter and s_j is the inverse scale parameter. Both parameters are positive real numbers. The reason that gamma distribution is selected as hyperpriors is that the gamma distribution is the conjugate prior to the likelihood.

The likelihood $P(Y_i^R|\sigma_{ii}^2,\sigma_{ii,0}^2,\eta_i)$ is actually the evidence separated according to each measurement in the first layer in Eq. 1. To get the expression of the evidence in the first layer, the joint probability (which is the product of likelihood and priors in the first layer) can be integrated over the unknown parameters x and δ. According to whether there is a gross error or not, the problem can be addressed in two different cases

$$P(Y_i^R|\alpha_i,\alpha_{i0},\eta_i)=P(Y_i^R|\sigma_{ii}^2,\sigma_{ii,0}^2,\eta_i)=\begin{cases}(2\pi)^{(-m+1)/2}(\alpha_i)^{(m-1)/2}m^{-1/2}\exp\left\{-\frac{\alpha_i}{2}\left(\sum_{j=1}^{m}y_{ij}^2-\frac{1}{m}\left(\sum_{j=1}^{m}y_{ij}\right)^2\right)\right\} & \text{if } \eta_i=1\\ (2\pi)^{-m/2}(\alpha_i)^{m/2}\alpha_{i0}^{1/2}(\alpha_{i0}+m\alpha_i)^{-1/2}\exp\left\{\left(\frac{\mu_{i0}^2\alpha_{i0}}{-2}+\frac{\alpha_i\sum_{j=1}^{m}y_{ij}^2}{-2}\right)-\frac{\left(\mu_{i0}\alpha_{i0}+\alpha_i\sum_{j=1}^{m}y_{ij}\right)^2}{-2(\alpha_{i0}+m\alpha_i)}\right\} & \text{if } \eta_i=0\end{cases} \tag{19}$$

Substituting Eqs. 18 and 19 into Eq. 15, the posterior probability of the hyperparameters can be written as

$$P(\alpha_i,\alpha_{i0}|\eta_i,Y_i^R)=P(\sigma_{ii}^2,\sigma_{ii,0}^2|\eta_i,Y_i^R)\propto P(Y_i^R|\sigma_{ii}^2,\sigma_{ii,0}^2,\eta_i)P(\sigma_{ii}^2,\sigma_{ii,0}^2|\eta_i)$$

$$=\begin{cases}\alpha_i^{(m-1)/2}\exp\left\{\dfrac{\sum_{j=1}^m y_{ij}^2-\frac{1}{m}(\sum_{j=1}^m y_{ij})^2}{-2}\alpha_i\right\}\alpha_i^{k_i-1}\exp(-s_i\alpha_i)\alpha_{i0}^{k_{i0}-1}\exp(-s_{i0}\alpha_{i0}) & \text{if } \eta_i=1\\ \alpha_i^{m/2}\alpha_{i0}^{1/2}(\alpha_{i0}+m\alpha_i)^{-1/2}\exp\left\{-\frac{1}{2}\left(\mu_{i0}^2\alpha_{i0}+\alpha_i\sum_{j=1}^m y_{ij}^2-\dfrac{(\mu_{i0}\alpha_{i0}+\alpha_i\sum_{j=1}^m y_{ij})^2}{\alpha_{i0}+m\alpha_i}\right)\right\} & \text{if } \eta_i=0\\ \quad\times\alpha_i^{k_i-1}\exp(-s_i\alpha_i)\alpha_{i0}^{k_{i0}-1}\exp(-s_{i0}\alpha_{i0})\end{cases} \tag{20}$$

To find the posterior mode, the above posterior distribution can be maximized, or equivalently, the negative logarithm of the posterior probability can be minimized. It leads to the following optimization problem

$$\max_{\alpha_i,\alpha_{i0}} P(\alpha_i,\alpha_{i0}|\eta_i,Y_i^R)=\min_{\alpha_i,\alpha_{i0}} J_2(\alpha_i,\alpha_{i0}) \tag{21}$$

where

$$J_2(\alpha_i,\alpha_{i0})$$

$$=\begin{cases}-\dfrac{m-1}{2}\log\alpha_i+\dfrac{\sum_{j=1}^m y_{ij}^2-\frac{1}{m}(\sum_{j=1}^m y_{ij})^2}{2}\alpha_i-(k_i-1)\log\alpha_i+s_i\alpha_i-(k_{i0}-1)\log\alpha_{i0}+s_{i0}\alpha_{i0} & \text{if } \eta_i=1\\ -\dfrac{m}{2}\log\alpha_i-\dfrac{1}{2}\log\alpha_{i0}+\dfrac{1}{2}\log(\alpha_{i0}+m\alpha_i)+\dfrac{1}{2}\left(\mu_{i0}^2\alpha_{i0}+\alpha_i\sum_{j=1}^m y_{ij}^2-\dfrac{\left(\mu_{i0}\alpha_{i0}+\alpha_i\sum_{j=1}^m y_{ij}\right)^2}{\alpha_{i0}+m\alpha_i}\right) & \text{if } \eta_i=0\\ \quad -(k_i-1)\log\alpha_i+s_i\alpha_i-(k_{i0}-1)\log\alpha_{i0}+s_{i0}\alpha_{i0}\end{cases} \tag{22}$$

As α_i and α_{i0} are positive variables, we can take logarithm on them. The gradient of the cost function $J_2(\alpha_i,\alpha_{i0})$ is

$$\frac{\partial J_2(\alpha_i,\alpha_{i0})}{\partial\log\alpha_i}$$

$$=\begin{cases}-\dfrac{m}{2}+\dfrac{1}{2}+\dfrac{1}{2}\left[\sum_{j=1}^m {y_{ij}}^2-\frac{1}{m}(\sum_{j=1}^m y_{ij})^2\right]\alpha_i-(k_i-1)+s_i\alpha_i & \text{if } \eta_i=1\\ -\dfrac{m}{2}+\dfrac{m\alpha_i}{2(\alpha_{i0}+m\alpha_i)}+\dfrac{1}{2}\sum_{j=1}^m {y_{ij}}^2\alpha_i-(k_i-1)+s_i\alpha_i & \text{if } \eta_i=0\\ -\dfrac{2(\mu_{i0}\alpha_{i0}+\sum_{j=1}^m y_{ij}\alpha_i)(\alpha_{i0}+m\alpha_i)\sum_{j=1}^m y_{ij}\alpha_i-(\mu_{i0}\alpha_{i0}+\sum_{j=1}^m y_{ij}\alpha_i)^2 m\alpha_i}{2(\alpha_{i0}+m\alpha_i)^2}\end{cases} \tag{23}$$

$$\frac{\partial J_2(\alpha_i,\alpha_{i0})}{\partial \log\alpha_{i0}} = \begin{cases} -(k_{i0}-1)+s_{i0}\alpha_{i0} \quad \text{if} \quad \eta_i=1 \\ -\frac{1}{2}+\frac{\alpha_{i0}}{2(\alpha_{i0}+m\alpha_i)}+\frac{1}{2}{\mu_{i0}}^2\alpha_{i0}-(k_{i0}-1)+s_{i0}\alpha_{i0} \quad \text{if} \quad \eta_i=0 \\ -\frac{2(\mu_{i0}\alpha_{i0}+\sum_{j=1}^{m}y_{ij}\alpha_i)(\alpha_{i0}+m\alpha_i)\mu_{i0}\alpha_{i0}-(\mu_{i0}\alpha_{i0}+\sum_{j=1}^{m}y_{ij}\alpha_i)^2\alpha_{i0}}{2(\alpha_{i0}+m\alpha_i)^2} \end{cases} \tag{24}$$

As it is difficult to get the analytical solution of α_i and α_{i0}, numerical optimization method (e.g., Newton's method, etc.) is used to find the solution of Eq. 21. The variance σ_{ii}^2 and $\sigma_{ii,0}^2$ can be obtained by taking inverse of α_i and α_{i0}, respectively. After calculating σ_{ii}^2 and $\sigma_{ii,0}^2$ for each measurement, the covariance matrix Σ and Σ_0 can be constructed.

The posterior distribution given by Eq. 20 is complex in general so that it cannot be directly applied in the three-layered optimization framework. An approximation of the posterior distribution is a key to simplify the calculation in the hierarchical Bayesian approach. MacKay[21] introduced a method called Laplace Approximation which can approximate posterior using a normal distribution. To approximate the posterior distribution by a normal distribution, we approximate the logarithm of the posterior distribution by its second-order Taylor expansion around the most probable estimation, $\Theta^{\mathrm{MP}}=(log\alpha_i^{\mathrm{MP}}, log\alpha_{i0}^{\mathrm{MP}})$, and the posterior distribution can be approximated by

$$\begin{aligned} \log P(\log\alpha_i,\log\alpha_{i0}|\eta_i,Y_i^R) \approx{}& \log P(\log\alpha_i,\log\alpha_{i0}|\eta_i,Y_i^R)|_{\log\alpha_i^{\mathrm{MP}},\log\alpha_{i0}^{\mathrm{MP}}} \\ &+\nabla\log P(\log\alpha_i,\log\alpha_{i0}|\eta_i,Y_i^R)|_{\log\alpha_i^{\mathrm{MP}},\log\alpha_{i0}^{\mathrm{MP}}} \\ &+\frac{1}{2}z_i^T\nabla\nabla\log P(\log\alpha_i,\log\alpha_{i0}|\eta_i,Y_i^R)|_{\log\alpha_i^{\mathrm{MP}},\log\alpha_{i0}^{\mathrm{MP}}}z_i \end{aligned} \tag{25}$$

where $z_i=[\log\alpha_i-\log\alpha_i^{\mathrm{MP}},\log\alpha_{i0}-\log\alpha_{i0}^{\mathrm{MP}}]^T$. Because $\log\alpha_i^{\mathrm{MP}}$ and $\log\alpha_{i0}^{\mathrm{MP}}$ correspond to a maximum of the logarithm of the posterior distribution, the second term on the right-hand side of Eq. 25 equals to zero. Using Laplace Approximation, the posterior distribution can be approximated by a normal distribution as following

$$\begin{aligned} P(\log\alpha_i,\log\alpha_{i0}|\eta_i,Y_i^R) &\approx P(\log\alpha_i^{\mathrm{MP}},\log\alpha_{i0}^{\mathrm{MP}}|\eta_i,Y_i^R)\exp(-\frac{1}{2}{z_i}^TQ_iz_i) \\ &=(2\pi)^{-1/2}\sqrt{\det Q_i}\times\exp(-\frac{1}{2}z_i^TQ_iz_i) \end{aligned} \tag{26}$$

where Q_i is the Hessian matrix of the cost function J_2 evaluated at $log\alpha_i^{\mathrm{MP}}$ and $log\alpha_{i0}^{\mathrm{MP}}$

$$Q_i=-\nabla\nabla\log P(\log\alpha_i,\log\alpha_{i0}|\eta,D)|_{log\alpha_i{}^{\mathrm{MP}},log\alpha_{i0}{}^{\mathrm{MP}}} \tag{27}$$

The Hessian matrix Q_i can be calculated as

$$Q_i=\begin{pmatrix} Q_{i,11} & Q_{i,12} \\ Q_{i,21} & Q_{i,22} \end{pmatrix}=\begin{pmatrix} \frac{\partial^2 J_2}{\partial\log\alpha_i{}^2} & \frac{\partial^2 J_2}{\partial\log\alpha_i\partial\log\alpha_{i0}} \\ \frac{\partial^2 J_2}{\partial\log\alpha_{i0}\partial\log\alpha_i} & \frac{\partial^2 J_2}{\partial\log\alpha_{i0}{}^2} \end{pmatrix} \tag{28}$$

Third layer: Inference of the indicator η_i

In the above derivations, it is assumed that the gross error indicators are known. In the third layer, the objective is to identify the indicators. Applying Bayes rule and separating the posterior corresponding to each measurement, the posterior distribution is obtained as following

$$P(\eta_i|Y_i^R)=\frac{P(Y_i^R|\eta_i)P(\eta_i)}{P(Y_i^R)}\propto P(Y_i^R|\eta_i)P(\eta_i) \tag{29}$$

For priors, it is assumed that each instrument has a probability p_i to contain a gross error, then the probability of η_i given p_i can be expressed as

$$P(\eta_i|p_i)={p_i}^{\eta_i}(1-p_i)^{1-\eta_i} \tag{30}$$

and the prior for p_i is set to be a Beta distribution

$$P(p_i)=\frac{\Gamma(r_i+\beta_i)}{\Gamma(r_i)\Gamma(\beta_i)}{p_i}^{r_i-1}(1-p_i)^{\beta_i-1} \tag{31}$$

Then, by integrating out p_i, we can get the prior distribution for η_i

$$\begin{aligned} P(\eta_i)&=\int_0^1 P(\eta_i|p_i)P(p_i)dp_i \\ &=\int_0^1\frac{\Gamma(r_i+\beta_i)}{\Gamma(r_i)\Gamma(\beta_i)}{p_i}^{\eta_i+r_i-1}(1-p_i)^{1-\eta_i+\beta_i-1}dp_i \\ &=\frac{\Gamma(r_i+\beta_i)}{\Gamma(r_i)\Gamma(\beta_i)}\frac{\Gamma(r_i+\eta_i)\Gamma(1-\eta_i+\beta_i)}{\Gamma(r_i+\beta_i+1)} \end{aligned} \tag{32}$$

The likelihood in the third layer is equal to the evidence in the second layer in Eq. 15. As we divide the problem into two cases according to whether there is a gross error or not in the second layer, the likelihood in the third layer can be handled in two cases by substituting Eqs. 18, 19, and 26 in Eq. 15. However, for simplicity, we write the likelihood in a general form as following

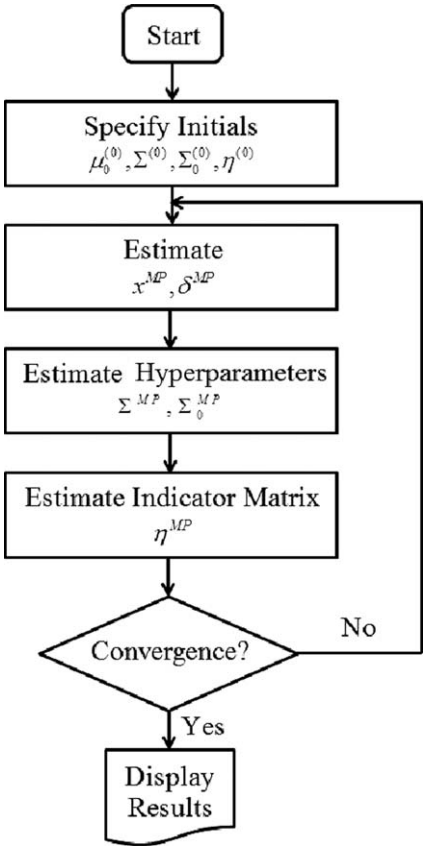

Figure 1. Flow chart of Algorithm 1.

Algorithm 1: Bayesian Method for Simultaneous Gross Error Detection and Data Reconciliation

input: Sample data D, and selected initials for indicator matrix η

initialize: Variance of the random error:
$\sigma_{ii}^{2(0)} = \text{MAD}/0.6745$,
hyperparameters: a large value for $\sigma_{ii,0}^{2(0)}$ and the reconciled data for $\mu_0^{(0)}$

while the estimations of the parameters do not change within tolerance **do**

Step 1. In the first layer, maximize $P(x^{(k)}, \delta^{(k)}|D, \Sigma^{(k-1)}, \Sigma_0^{(k-1)}, \eta^{(k-1)})$ to update the MAP estimates of x and δ, $x^{(k)}$ and $\delta^{(k)}$ using Eqs. 12 and 13.

Step 2. Update $\mu_0^{(k)}$ as the estimates of x, i.e., $\mu_0^{(k)} = x^{(k)}$.

Step 3. In the second layer, maximize $P(\alpha_i^{(k)}, \alpha_{i0}^{(k)}|\eta_i^{(k-1)}, Y_i^R)$ for each measurement to update the MAP estimates of hyperparameters, $\alpha_i^{(k)}$ and $\alpha_{i0}^{(k)}$. The variance can be obtained by taking an inverse of α_i and α_{i0}.

Step 4. In the third layer, evaluate the posterior probability of each measurement, $P(\eta_i=1|Y_i^R)$ and $P(\eta_i=0|Y_i^R)$. Compare the values of them. If the value of $P(\eta_i=1|Y_i^R)$ is larger, set $\eta_i^{(k)}=1$, otherwise, set $\eta_i^{(k)}=0$.

Step 5. Update the estimated values of $\Sigma^{(k)}$, $\Sigma_0^{(k)}$ and $\eta^{(k)}$ and use them in the next iteration.

end while

return: Estimations of different parameters $\hat{x}$, $\hat{\delta}$, $\hat{\Sigma}$, $\hat{\Sigma_0}$ and $\hat{\eta}$

$$P(Y_i^R|\eta_i) = \frac{P(Y_i^R|\alpha_i, \alpha_{i0}, \eta_i)P(\alpha_i, \alpha_{i0}|\eta_i)}{P(\log\alpha_i, \log\alpha_{i0}|\eta_i, Y_i^R)} \tag{33}$$

Since the left-hand side is not relevant to α_i and α_{i0}, the right-hand side has no relationship with α_i and α_{i0}, which means that the values of α_i and α_{i0} have no influence on the value of the likelihood. The MAP of α_i and α_{i0} is taken to simplify the calculation

$$\begin{aligned} P(Y_i^R|\eta_i) &= \frac{P(Y_i^R|\alpha_i, \alpha_{i0}, \eta_i)P(\alpha_i, \alpha_{i0}|\eta_i)}{P(\log\alpha_i, \log\alpha_{i0}|\eta_i, Y_i^R)}\Big|_{\alpha_i^{\text{MP}}, \alpha_{i0}^{\text{MP}}} \\ &\approx \frac{P(Y_i^R|\alpha_i, \alpha_{i0}, \eta_i)P(\alpha_i, \alpha_{i0}|\eta_i)}{\sqrt{\det Q_i}}\Big|_{\alpha_i^{\text{MP}}, \alpha_{i0}^{\text{MP}}} \end{aligned} \tag{34}$$

After the above derivations, $P(\eta_i=1|Y_i^R)$ and $P(\eta_i=0|Y_i^R)$ are both calculated. They are further compared to determine the value of η_i. The indicator matrix η can be obtained after the value of the indicator for each measurement is decided.

Simultaneous gross error detection and data reconciliation procedure

The implementation procedure of the hierarchical Bayesian approach is outlined as Algorithm 1.

The principle of the initial value selection is discussed as following. $\sigma_{ii}^{2(0)} = \text{MAD}/0.6745$ is utilized to calculate the robust variance. This equation is commonly used in robust regression. The constant 0.6745 makes the estimate of σ_{ii}^2 unbiased for the normal distribution. MAD is the median absolute deviation of the residuals; a large value is set for $\sigma_{ii,0}^{2(0)}$ to reduce the influence of not setting an accurate initial value for $\mu_{i0}^{(0)}$. μ_0 is the mean value in the hyperprior for x.

In Algorithm 1, Step 1 to 5 will be repeated iteratively until no further improvements (the estimations of the parameters do not change) are gained. The flow chart of Algorithm 1 is shown in Figure 1.

Although Algorithm 1 can be directly applied for gross error identification, it may lead to too many mispredictions. To improve the performance of the proposed simultaneous gross error detection and data reconciliation approach, a serial strategy is introduced in the next section.

Serial Strategy for Improved Performance

Algorithm 1 may lead to many mispredictions, which is mainly due to the smearing effect.[1] All the variables are related through the constraints according to the network of the system. A gross error in one measurement may influence other measurements, so that the probability $P(\eta_i=1|Y_i^R)$ can be larger than $P(\eta_i=0|Y_i^R)$ for a process measurement without gross error. There are many factors influencing the degree of smearing, such as the level of redundancy, the size of the standard deviations of the random error, and the magnitudes of the gross errors.[1,29]

Furthermore, simultaneous gross error estimation and data reconciliation cannot be solved for an arbitrary set of GEC due to the existence of equivalent sets.[30] The concept of equivalent sets is defined as follows: if two sets of gross errors have the same effect in data reconciliation, they are regarded as equivalent sets, which means that when simulating either one, they both lead to the same objective function value of data reconciliation. Equivalent sets exist when candidate streams are

in the same loop in an augmented graph consisting of the original graph representing the flow sheet with the addition of environmental node. The environment node is an additional node of the flow sheet so that all process feeds and products can connect to it. It is impossible to distinguish the equivalent sets and this leads to the conditions of exact detectability in simultaneous gross error detection and data reconciliation problem. There are mainly two rules for this detectability problem[30]

1. The maximum number of gross errors that can be simulated in an open system is equal to the number of process units (blocks in the diagram).

2. A set of gross errors can be exactly detected only if no subset of these variables forms a loop with an additional stream. In other words, a set of gross errors can be exactly detected only if the corresponding set of columns of the incidence matrix A does not form a linearly dependent set with any other additional column.

To address the above issues and improve the performance of Algorithm 1, a serial strategy is adopted. The serial strategy identifies gross errors in a sequential mode. This strategy is based on the assumption that for any system, the probability of containing k gross errors is larger than the probability of containing $k+1$ gross errors. This strategy can detect the minimum possible number of gross errors and solve data reconciliation problem and estimation problem with all suspects. The strategy is outlined as Algorithm 2 the flow chart of Algorithm 2 is shown in Figure 2.

Algorithm 2: Serial Strategy for Simultaneous Multiple Gross Error Detection and Data Reconciliation

Input: Sample data D

Initialize: Create two lists: one is for gross error candidates (GEC) and another is for the final gross errors (FGE). Set them empty at first. Set the initial values for η_i to be all 1*s*

Step 1. **Run Algorithm** 1 once to get the gross error candidates.
if there are r variables in suspect ($r > 0$)
go to step 2.
else
declare no gross error and stop.
end if-else

Step 2. Put all r variables detected in step 1 in GEC.
if GEC and FGE have same elements
erase them in GEC.
end if

while GEC is not empty **do**

Step 3. Run Algorithm 1 with the initial values to be all the members in FGE and one member of the GEC at a time. If there are r variables in GEC, then we need to run Algorithm 1 r times.

Step 4. Calculate the ratio $\frac{P(\eta_i=1|Y_i^R)}{P(\eta_i=0|Y_i^R)}$. For each run find the value of the ratio corresponding to the simulated member in GEC. Compare these values and determine which member of the GEC leads to the largest value of ratio and get the corresponding gross error candidates. Add that variable to the FGE. Empty set GEC.

Step 5. Add the gross error candidates obtained in step 4 into the GEC. If any members in the GEC are in the same loop with the members in the FGE, erase the member in the GEC with the smallest ratio and also eliminate the member which is already in the FGE.

end while

Step 6. Determine all equivalent sets.

return: Estimations of different parameters x, δ, Σ, Σ_0 and η. If there are equivalent sets for the gross errors, the equivalent sets are also returned.

Simulation Study

Linear case

In this section, simulated data sets are used to study the effectiveness of the proposed Bayesian approach for simultaneous gross error detection and data reconciliation. The process network shown in Figure 3 is considered. There are four units and seven flow rates.

From the above process network, the flow rates satisfy the following material balance

$$x_1 - x_2 + x_4 = 0$$
$$x_2 - x_3 + x_6 = 0$$
$$x_3 - x_4 - x_5 = 0$$
$$x_5 - x_6 - x_7 = 0$$

All the variables are assumed to be measured and the above equations can be further written as $Ax = 0$, where $x = [x_1, x_2, ..., x_7]^T$ and the incidence matrix A is

$$A = \begin{bmatrix} 1 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 \end{bmatrix}$$

The data are generated randomly using MATLAB. The vector of the correct values of the flow rates is set as $x = [1 \quad 2 \quad 3 \quad 1 \quad 2 \quad 1 \quad 1]^T$. To avoid confusion between the gross errors and the noise (random error), the value of the covariance matrix Σ cannot be too large.

Case 1: Single Gross Error. In this case, only one gross error is introduced in the first measurement x_1 and the magnitude of the gross error is set to $\delta_1 = 2$. As there are no gross errors in other measurements, the magnitudes of other gross errors are set to zeros.

We compare the results with two different standard deviations (variance) of the measurements:

1. $\Sigma = 0.0016I$, where I is identity matrix, the standard deviations are around 1.3–4% of the measurements.

2. $\Sigma = 0.1I$, the standard deviations are around 11–32% of the measurements.

A set of simulated measurement data with the two different standard deviations are plotted in Figures 4 and 5, respectively.

To evaluate the performance of the proposed method, 50 sets of simulated measurement data are generated and used for gross error detection and data reconciliation. With Algorithm 2, Figures 6–8 show the histograms of the estimated value of x, the estimated value of δ and the estimated value of the standard deviation σ with $\Sigma = 0.0016I$, respectively. Figures 9–11 show the histograms of the estimated value of x, the estimated value of δ (for both cases $\delta_1 = 2$, $\delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta_6 = \delta_7 = 0$) and the

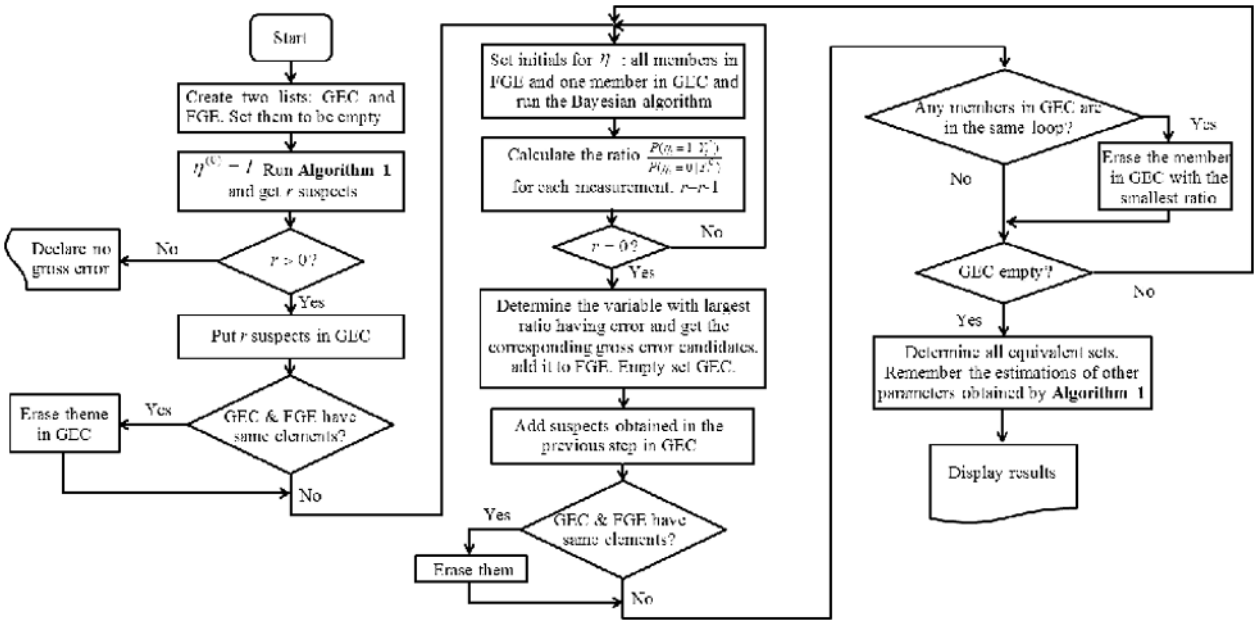

Figure 2. Flow chart of Algorithm 2.

estimated value of the standard deviation σ with $\Sigma=0.1I$, respectively.

The performances of Algorithm 1 and Algorithm 2 are compared for all the cases of one gross error introduced. For each case, one of the seven measurements contains one gross error and the simulation is performed for 50 runs using Algorithm 1 and Algorithm 2, respectively.

A performance measure—correct rate is defined for performance evaluation

$$\text{Correct Rate} = \frac{\text{Number of Runs that All the Gross Errors are Correctly Identified}}{\text{Total Number of Runs}}$$

Table 1 displays the correct rates of Algorithm 1 and Algorithm 2. The results with different standard deviations of measurements are also compared.

It can be seen from Table 1 that Algorithm 1 has more mispredictions than Algorithm 2. Since in this case only single gross error is considered, according to the flow sheet there is no equivalent set for each case. The mispredictions may be caused by smearing effect. The confusion between the random error and the gross error may lead to the mispredictions (the magnitude δ of the gross error is only slightly larger than the 3σ). Algorithm 2 enhances the correct rate of gross error detection and estimation. It is worth mentioning that the existing methods generally handle small standard deviations. For example, the standard deviations of the measurements are chosen to be 2% of each measurement by Bagajewicz and Jiang.[31] The proposed method in this work can deal with much larger standard deviations. The proposed approach also works well with small standard deviation as expected. Algorithm 2 leads to 100% correct rate for $\Sigma=0.0016I$. Even though the standard deviations are increased to 13–32% of the measurements, the correct rate shows that Algorithm 2 is still efficient to detect the gross error. As the proposed method can handle large standard deviation in the data, only $\Sigma=0.1I$ is considered in the rest of the case studies.

Case 2: Multiple Gross Errors without Equivalent Sets. In this case, multiple gross errors in the data are investigated. We first study the case that no equivalent set exists in the gross errors. Algorithm 2 is illustrated using the simulation example under different scenarios. Tables 2 and 3 below show the results in each steps following the procedure of the serial strategy. The gross error initial values for Algorithm 1 are shown in the second column of the tables. The third column of the tables provides the estimations of the magnitudes of the gross errors. The values of the ratio $\frac{P(\eta_i=1|Y_i^R)}{P(\eta_i=0|Y_i^R)}$ are listed in the fourth column. The last column shows the variables as GEC flagged by Algorithm 1.

Table 2 shows the solution procedure for two gross errors in x_2 and x_7 with magnitudes $\delta_2=3$ and $\delta_7=1$, respectively. In the first step, the initial values for indicator η_i are set to be all 1*s*, and the gross errors candidates flagged by Algorithm 1 are 2 and 6. In the second step, the flagged GEC (2 and 6) are put into GEC. In the third step, Algorithm 1 is run twice as there are two elements in the GEC. The initial values of indicator η_i are set as $\eta_2=1$ (others are set to zero) for the first run and $\eta_6=1$ (others are set to be zero) for the second run. In the fourth step, we calculate the ratio $\frac{P(\eta_2=1|Y_2^R)}{P(\eta_2=0|Y_2^R)}$ and $\frac{P(\eta_6=1|Y_6^R)}{P(\eta_6=0|Y_6^R)}$ for the two runs in step 3 and compare them. The run with initial $\eta_2=1$ has a larger ratio, so element 2 in GEC is identified as a variable with gross error (i.e., the second measurement is detected to contain a gross error). Element 2 is added to the FGE. The GEC is set empty. In the fifth step, we get the GEC corresponding to element 2, which are 2 and 7 and put them in the GEC. Since 2 is already in the FGE, it is eliminated in the GEC and GEC has only one element 7 at this time. The GEC is not empty so the algorithm goes to step 3; we run Algorithm 1 with initials to be all the members in

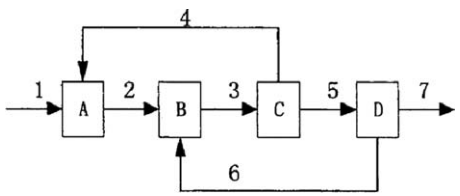

Figure 3. Diagram of process network.

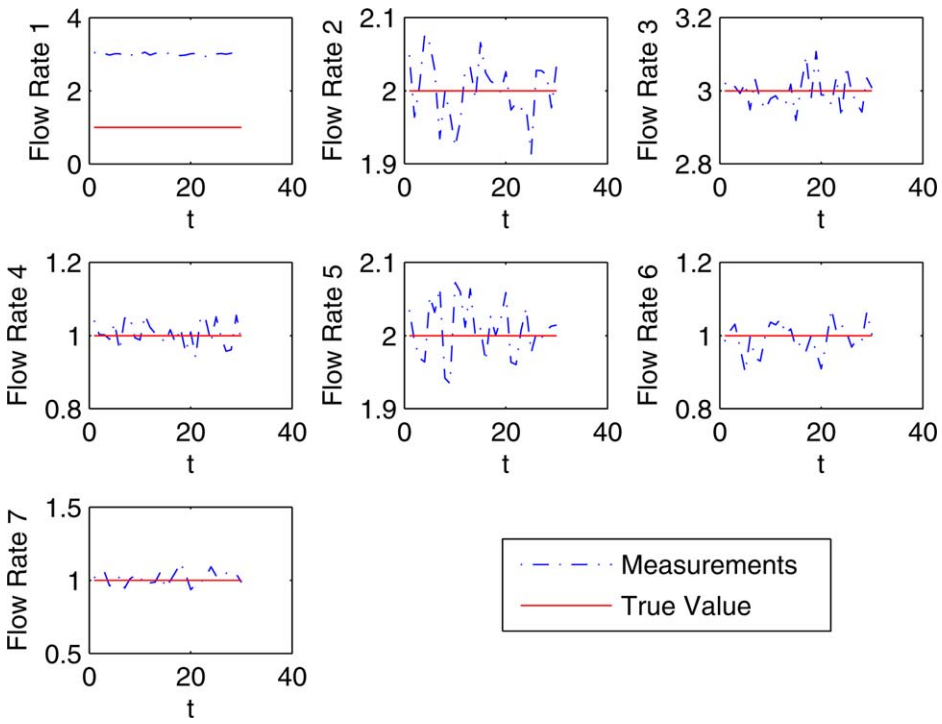

Figure 4. Data plot with gross error in the first measurement, $\Sigma=0.0016I$.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the FGE (2) and one member of GEC (7). The GEC flagged by Algorithm 1 are 2 and 7. In the second step 5, the GEC is empty and the algorithm is stopped. Even though the gross error in the seventh measurement is not flagged by the Algorithm 1 in step 1, the gross errors are still successfully identified. The number of iterations that is needed to detect all the gross errors in Algorithm 2 is 2 which is the same as the number of gross error introduced. Using Algorithm 2, the gross errors are exactly detected. Then, the magnitudes of the gross errors can be estimated. After the gross errors are successfully detected and the magnitudes of the gross errors are estimated, the correct values of x and the covariance matrix Σ of the random errors can be estimated.

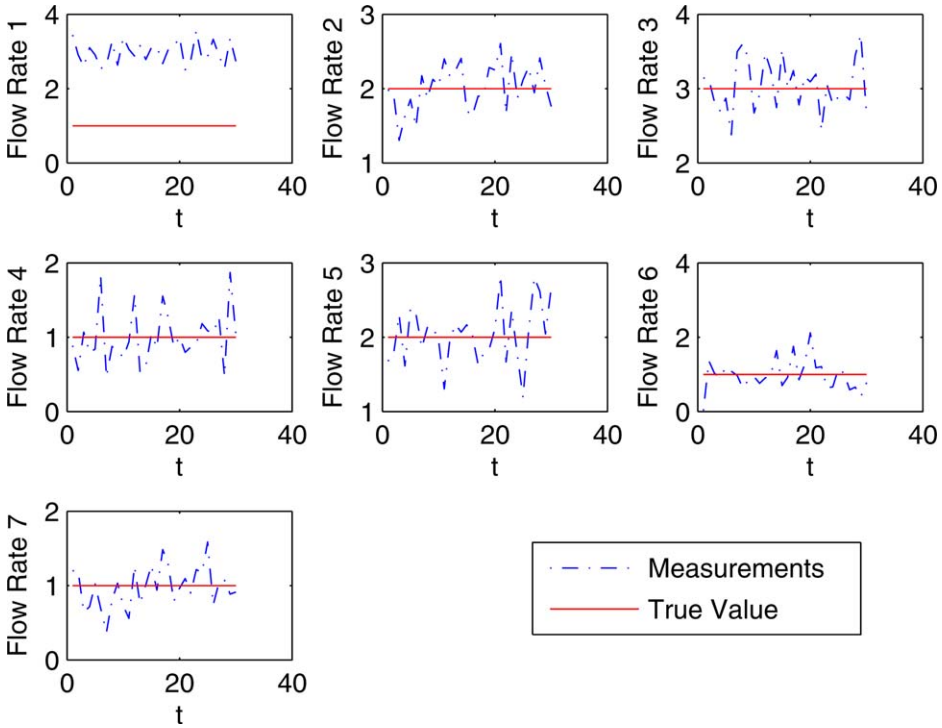

Figure 5. Data plot with gross error in the first measurement, $\Sigma=0.1I$.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

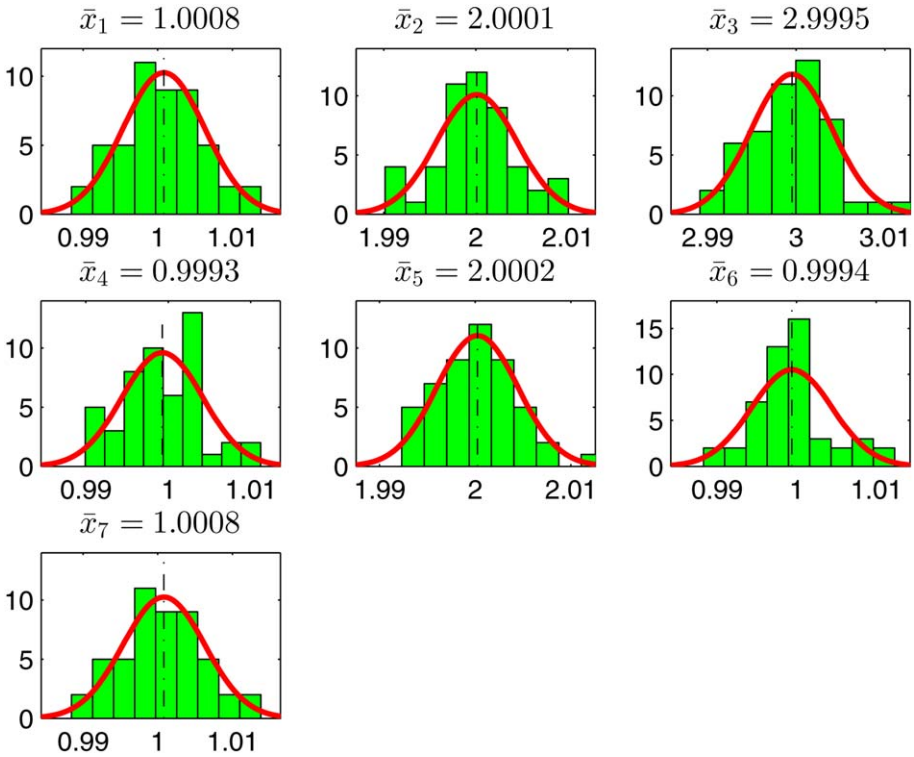

Figure 6. Histogram of the estimated value of x, $\Sigma=0.0016I$ (50 runs).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The final estimated values are

$$\hat{x}=[\,0.9749 \quad 2.0105 \quad 3.0604 \quad 1.0356 \quad 2.0248 \quad 1.0499 \quad 0.9749\,]^T$$

$$\hat{\delta}_2=3.0,\ \hat{\delta}_7=1.0,\ \hat{\delta}_1=\hat{\delta}_3=\hat{\delta}_4=\hat{\delta}_5=\hat{\delta}_6=0$$

$$\hat{\sigma}^2_{11}=0.0918,\ \hat{\sigma}^2_{22}=0.1085,\ \hat{\sigma}^2_{33}=0.1109,\ \hat{\sigma}^2_{44}=0.1117,$$
$$\hat{\sigma}^2_{55}=0.0919,\ \hat{\sigma}^2_{66}=0.0878$$

$$\hat{\sigma}^2_{77}=0.0845$$

In Table 3, three biases are considered with $\delta_1=2$, $\delta_2=3$, and $\delta_5=4$. The procedure is similar to Table 2. As the number of gross errors introduced is larger than the above cases, Algorithm 2 takes 1 more iteration to identify all the gross errors successfully.

The final estimated values are

$$\hat{x}=[\,1.0092 \quad 1.9142 \quad 2.9036 \quad 0.9050 \quad 1.9986 \quad 0.9894 \quad 1.0092\,]^T$$

$$\hat{\delta}_1=1.9,\ \hat{\delta}_2=3.1,\ \hat{\delta}_5=3.9,\ \hat{\delta}_3=\hat{\delta}_4=\hat{\delta}_6=\hat{\delta}_7=0$$

$$\hat{\sigma}^2_{11}=0.0721,\ \hat{\sigma}^2_{22}=0.1150,\ \hat{\sigma}^2_{33}=0.0935,\ \hat{\sigma}^2_{44}=0.0838,$$
$$\hat{\sigma}^2_{55}=0.0973,\ \hat{\sigma}^2_{66}=0.1180$$

$$\hat{\sigma}^2_{77}=0.1168$$

In Tables 2 and 3, the number of the gross errors introduced is less than the number of the units in the process network and none of them form a loop with an additional variable, which means that the gross errors introduced satisfy the condition of exact detectability.

Case 3: Multiple Gross Errors with Equivalent Sets. In this case, multiple gross errors are investigated under the situation that equivalent sets exist. Here, two biases are introduced as $\delta_2=3$ and $\delta_3=4$. In the first step, the initial values for indicator η_i are set to be all 1s and the gross errors candidates flagged by Algorithm 1 are 3 and 4. Since there are two suspects, it proceeds to step 2. In the second step, the flagged GEC 3 and 4 are put into GEC. In the third step, Algorithm 1 is run twice since there are two elements in the GEC. The initial values of indicator η_i are different for different runs, that is, $\eta_3=1$ (others are set to be zero) for the first run and $\eta_4=1$ (others are set to be zero) for the second run. In the fourth step, we calculate the ratio $\frac{P(\eta_3=1|Y_3^R)}{P(\eta_3=0|Y_3^R)}$ and $\frac{P(\eta_4=1|Y_4^R)}{P(\eta_4=0|Y_4^R)}$ for the two runs in step 3 and compare them. The run with initial $\eta_4=1$

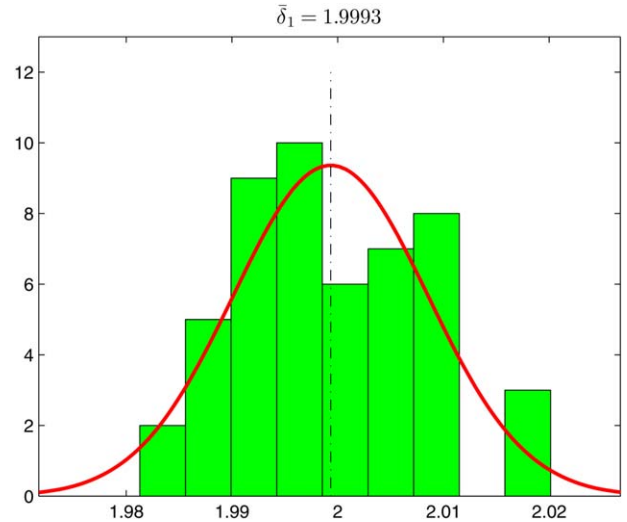

Figure 7. Histogram of the estimated value of δ_1, $\Sigma=0.0016I$ (50 runs).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

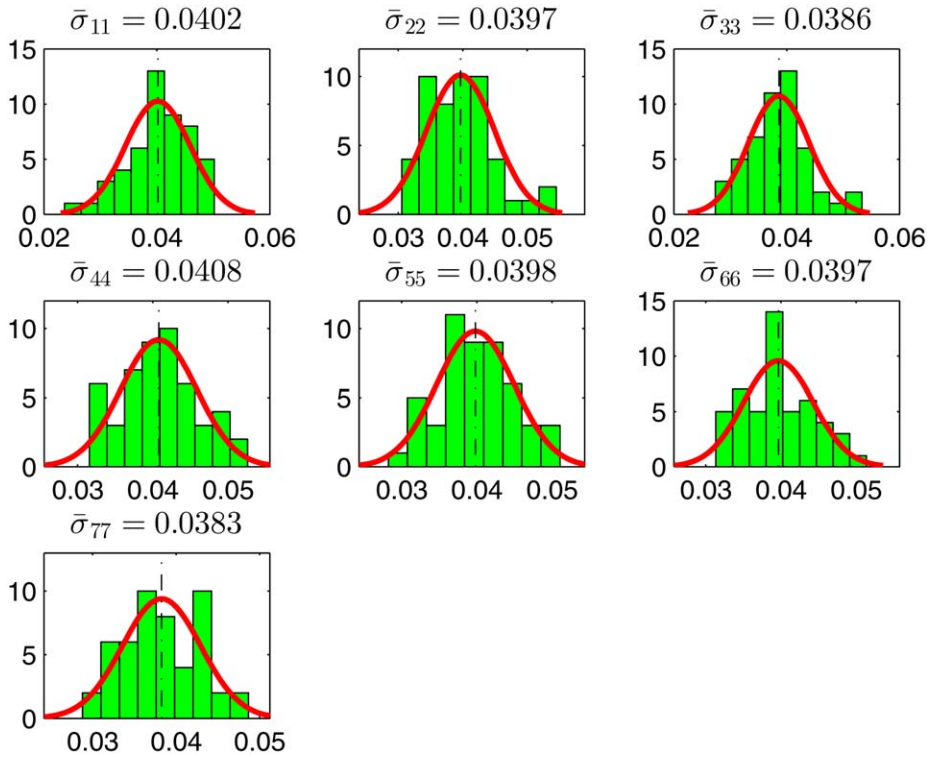

Figure 8. Histogram of the estimated value of σ, Σ=0.0016*I* (50 runs).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

has the largest ratio, so element 4 in GEC is determined to be the final gross errors (FGE), which means that the fourth measurement is detected to contain a gross error. Element 4 is added in the FGE. The GEC is set empty. In the fifth step, we get the GEC corresponding to element 4, which are 3 and 4 and put them in the GEC. Since 4 is already in the FGE, it is eliminated in the GEC and GEC has only one element 3 at this time. The GEC is not empty so the algorithm goes to step 3; we run Algorithm 1 with initials to be all the members in the FGE (4) and one member of GEC (3). The GEC flagged by

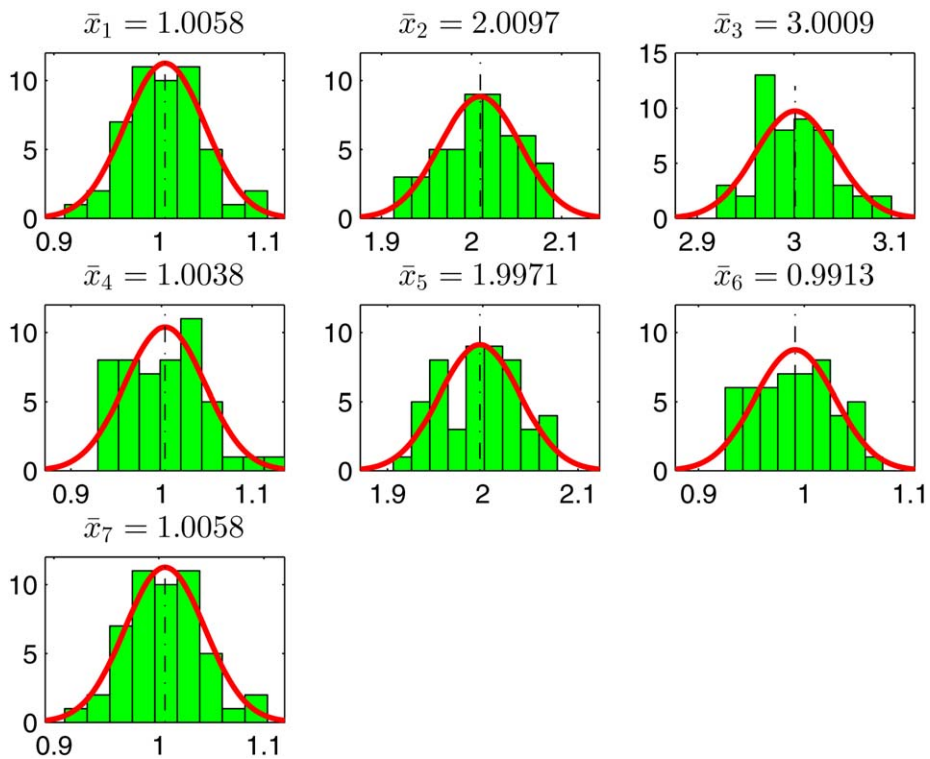

Figure 9. Histogram of estimated value of *x*, Σ=0.1*I* (50 runs).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

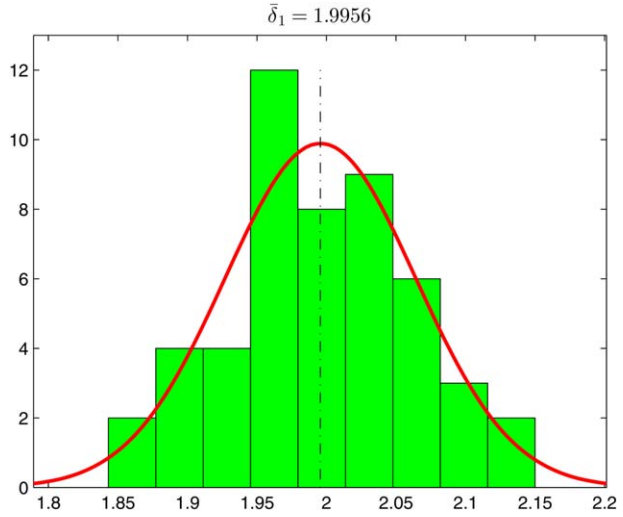

Figure 10. Histogram of the estimated value of δ_1, $\Sigma = 0.1I$ (50 runs).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Algorithm 1 are 3 and 4. In the second step 5, the GEC is empty and the algorithm is stopped. Since the second measurement and the third measurement form a loop with an additional stream 4, the gross errors introduced violate the second condition of exact detectability. It is difficult to identify the gross errors exactly in this case, but the equivalent sets are detected, as shown in Table 4.

Nonlinear case

The nonlinear programming problem in the first layer for both examples is solved in GAMS with solver CONOPT.

Case 1: Example from Pai and Fisher. This problem consists of eight variables, five of them are measured and three of them are unmeasured.[32] There are six nonlinear constraints which are shown as follows

$$0.5x_1^2 - 0.7x_2 + x_3u_1 + x_2^2u_1u_2 + 2x_3u_3^2 - 255.8 = 0$$

$$x_1 - 2x_2 + 3x_1x_3 - 2x_2u_1 - x_2u_2u_3 + 111.2 = 0$$

$$x_3u_1 - x_1 + 3x_2 + x_1u_2 - x_3\sqrt{u_3} - 33.57 = 0$$

$$x_4 - x_1 - x_3^2 + u_2 + 3u_3 = 0$$

$$x_5 - 2x_3u_2u_3 = 0$$

$$2x_1 + x_2x_3u_1 + u_2 - u_3 - 126.6 = 0$$

The exact values of the eight variables are

$$x_{\text{exact}} = [4.5124, 5, 5819, 1.9260, 1.4560, 4.8545]^T$$

$$u_{\text{exact}} = [11.070, 0.61467, 2.0504]^T$$

In this example, we attempt to demonstrate the effectiveness of the proposed simultaneous gross error detection and data reconciliation algorithm (Algorithm 1) and compare the results with the results of Tjoa and Biegler.[16] Following the same procedure provided by Tjoa and Biegle, the data are generated with the mean equal to the exact values and Gaussian noise level is 0.1. The algorithm is tested for 100 runs. Twenty percent of all measured variables are assumed to be shifted and the effect of location of gross errors is tested in three different situations. In situation 1, 5 blocks with 20 gross errors in each block are added in sequence from x_1 to x_5. In situation 2, all the measured variables are added with gross errors in every fifth run and no gross errors in other runs. In situation 3, one gross error is added for each run in sequence from x_1 to x_5 in

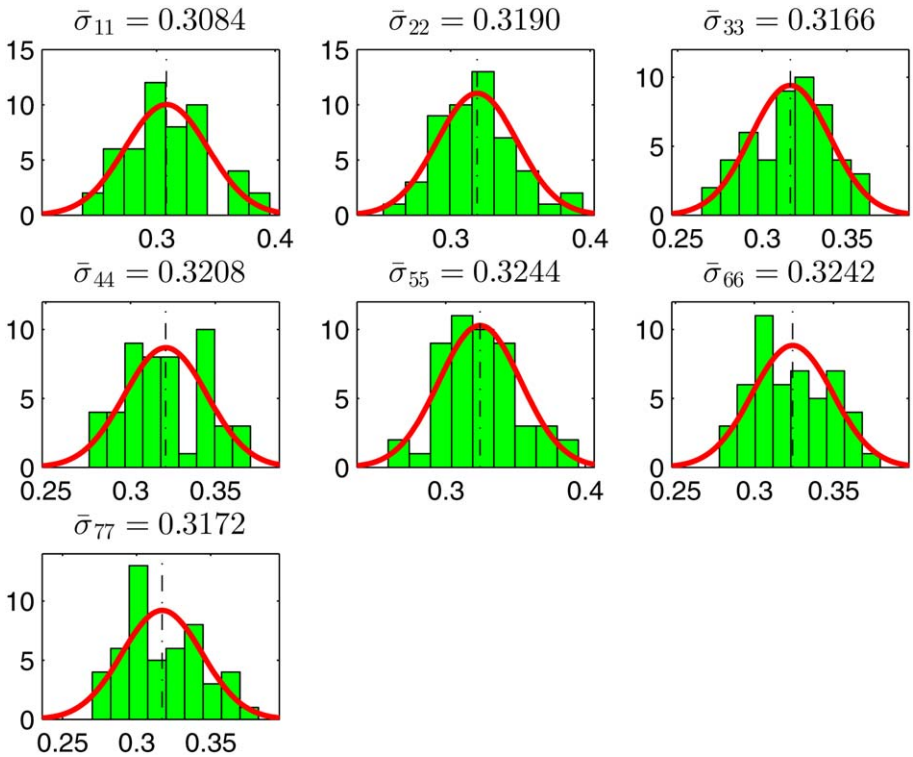

Figure 11. Histogram of the estimated value of σ, $\Sigma = 0.1I$ (50 runs).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 1. Comparison of Correct Rates (50 Runs)

Measurement Number	Correct Rate (Algorithm 1, $\Sigma=0.1I$)	Correct Rate (Algorithm 2, $\Sigma=0.0016I$)	Correct Rate (Algorithm 2, $\Sigma=0.1I$)
1	48/50 = 96%	50/50 = 100%	50/50 = 100%
2	46/50 = 92%	50/50 = 100%	49/50 = 98%
3	23/50 = 46%	50/50 = 100%	48/50 = 96%
4	35/50 = 70%	50/50 = 100%	50/50 = 100%
5	46/50 = 92%	50/50 = 100%	49/50 = 98%
6	34/50 = 68%	50/50 = 100%	48/50 = 96%
7	49/50 = 98%	50/50 = 100%	50/50 = 100%

Table 2. Algorithm 2 with Gross Error in x_2 (Magnitude: 3) and x_7 (Magnitude: 1)

Step Number	Gross Error Initials	Magnitude of Gross Error Estimated	Ratio	Variables Flagged by Algorithm 1
Step 1	All			2,6
Step 2	GEC: 2, 6			FGE: Empty
Step 3	2		1.8×10^{17}	2,7
	6		3.2×10^{3}	1,2,3,4,5,6
Step 4	FGE: 2			
Step 5	GEC: 7			
Step 3	2,7		46.8	2,7
Step 4	FGE: 2,7	2 (Magnitude: 3.0)		
		7 (Magnitude: 1.0)		
Step 5	GEC: Empty			
Step 6	Equivalent sets: none			

Table 3. Algorithm 2 with Gross Error in x_1 (Magnitude: 2), x_2(Magnitude: 3), and x_5 (Magnitude: 4)

Step Number	Gross Error Initials	Magnitude of Gross Error Estimated	Ratio	Variables Flagged by Algorithm 1
Step 1	All			2,3,5,7
Step 2	GEC: 2,3,5,7			FGE: Empty
Step 3	2		7.1×10^{13}	1,2,3,4,5,6,7
	3		8.8×10^{21}	2,3,4,5,6,7
	5		2.2×10^{23}	1,2,3,4,5,6,7
	7		8.1×10^{13}	1,2,3,4,5,6,7
Step 4	FGE: 5			
Step 5	GEC: 1,2,3,6,7			
Step 3	5,1		80.0	1,2,3,4,5,6,7
	5,2		1.3×10^{18}	1,2,5
	5,3		1.2×10^{10}	1,2,3,4,5,6,7
	5,6		1.3×10^{10}	1,2,4,5,6,7
	5,7		8.5×10^{6}	2,3,5,6,7
Step 4	FGE: 2,5			
Step 5	GEC: 1			
Step 3	1,2,5		8.9×10^{6}	1,2,5
Step 4	FGE: 1,2,5	1 (Magnitude: 1.9)		
		2 (Magnitude: 3.1)		
		5 (Magnitude: 3.9)		
Step 5	GEC: Empty			
Step 6	Equivalent sets: none			

Table 4. Algorithm 2 with Gross Error in x_2 (Magnitude: 3) and x_3 (Magnitude: 4)

Step Number	Gross Error Initials	Magnitude of Gross Error Simulated	Ratio	Variables Flagged by Algorithm 1
Step 1	All			3,4
Step 2	GEC: 3,4			FGE: Empty
Step 3	3		8.3×10^{9}	2,3,4
	4		7.6×10^{18}	3,4
Step 4	FGE: 4			
Step 5	GEC: 3			
Step 3	3,4		23.9	3,4
Step 4	FGE: 3,4	3 (Magnitude: 0.9)		
		4 (Magnitude: −3.2)		
Step 5	GEC: Empty			
Step 6	Equivalent sets: 2(3.2),3(4.1) and 2(−0.9),4(−4.1)			

Table 5. Results of Algorithm 1 for Gross Error Detection (δ=+0.4)

	x_1		x_2		x_3		x_4		x_5	
Situation	Right	Wrong	Right	Wrong	Right	Wrong	Right	Wrong	Right	Wrong
1 (Algorithm 1)	5	2	20	6	20	–	20	2	20	5
1 (Tjoa)	8	3	19	–	20	–	9	–	17	1
2 (Algorithm 1)	8	–	20	1	20	–	15	–	20	–
2 (Tjoa)	–	–	20	–	20	2	–	–	20	1
3 (Algorithm 1)	3	2	20	8	20	–	20	2	20	–
3 (Tjoa)	4	–	18	–	20	1	9	–	14	2

rotation. The magnitudes of gross errors are +0.4, +1, and +4, respectively.

For the case that the magnitude of gross errors is +0.4 (only 4σ), it is easy for the algorithm to be confused between the random error and gross error. The results are shown and compared in Table 5.

In Table 5, the number of a gross error correctly detected and the number of false alarm of a gross error for each variables are listed. From these numbers, we can see that the number of a gross error correctly detected obtained by Algorithm 1 is larger than Tjoa's method while the times of false alarm by Algorithm 1 is also larger than Tjoa's method. This is because that the correctly detected number of gross error and the false alarm are intimately related. If the probability of false alarm is allowed to be larger, the correct detected number of gross error will increase. For Algorithm 1, we can balance them by setting some hyperparameters, such as the hyperparameters for the prior distribution of p_i. However, the results show that Algorithm 1 is sufficient to detect the gross errors for the case that the magnitude of gross error is not easily distinguished from the random error.

Next, we consider the case when the magnitude of gross errors is set as $\delta = 1$. This is a more typical case of gross error detection and the results are shown in Table 6.

In Table 6, the number of a gross error correctly detected increases compared with the case of δ=0.4 while the number of false alarm decreases. This means that Algorithm 1 is more powerful for the case that the gross error can be distinguished from the random error. The performance of Algorithm 1 is very close to Tjoa's method. Situation 3 will be used for the studying of estimations of the magnitudes δ of gross errors, the estimations of the variance σ^2, and the estimations of the values of both measured and unmeasured variables (x and u). The estimated values listed below are the means for 100 runs.

For the case that $\delta=+1$, the final estimated values are

$$\hat{x}=[4.5354, 5.5749, 1.9231, 1.4963, 4.8236]^T$$

$$\hat{u}=[11.1733, 0.6200, 2.1344]^T$$

$$\hat{\delta}=[0.9717, 0.9969, 0.9995, 1.0000, 0.9931]^T$$

$$\hat{\sigma}_{11}^2=0.0103, \hat{\sigma}_{22}^2=0.0101, \hat{\sigma}_{33}^2=0.0107, \hat{\sigma}_{44}^2=0.0095, \hat{\sigma}_{55}^2=0.0100$$

For the case that $\delta=+4$, the final estimated values are

$$\hat{x}=[4.5439, 5.5539, 1.9228, 1.4094, 4.9022]^T$$

$$\hat{u}=[11.1428, 0.6128, 2.0728]^T$$

$$\hat{\delta}=[4.0102, 3.9862, 3.9975, 4.0028, 3.9951]^T$$

$$\hat{\sigma}_{11}^2=0.0102, \hat{\sigma}_{22}^2=0.0102, \hat{\sigma}_{33}^2=0.0100, \hat{\sigma}_{44}^2=0.0102, \hat{\sigma}_{55}^2=0.0104$$

It is shown that for all the cases, the estimated values are very close to the exact values. From the results shown above, we can see that the proposed algorithms work well for gross error detection and data reconciliation problem.If further improvement is needed for the problem, the serial strategy introduced above can be taken for exact detection of the gross error.

Furthermore, the performance of Algorithm 1 relies on the redundancy of the measurements. Since nonredundant variables can be adjusted only by the objective function not through the constraints. They may converge to the initial values (in the first layer). If the nonredundant variables contain gross errors, the reconciled data may be inaccurate and as the unmeasured variables are obtained through the constraints they may be influenced. The hyperparameter μ_0 is taken as the estimation of x and it is used in the third layer, so the performance of the gross error detection in the third layer may also be influenced. However, the redundancy of variables is related to both sensor network and the system constraints. Some existing methods can be applied to check the redundancy and we have added some related Refs. 33–35.

Case 2: Example from Swartz. The second example for nonlinear system is first described by Swartz for data reconciliation.[35] The flow sheet of the process is shown in Figure 12. Standard deviations of the measurements for flow rate and temperature are 2% and 0.75°, respectively. The model is more complex than the first example. There are 16 measured variables, 14 unmeasured variables, and 17 constraints. The constraints are obtained by considering mass and heat balances around the heat exchangers, splitting, and mixing junctions. To see the effectiveness of gross error detection, 10° is added to measurement TA7 as a gross error. Algorithm 1 can detect TA7 as a gross error and the final estimated values of both measured and unmeasured variables are shown in Table 7. The results are very close to the results obtained by Tjoa's method.[16] And, the computation time depends on the number

Table 6. Results of Algorithm 1 for Gross Error Detection (δ=+1)

	x_1		x_2		x_3		x_4		x_5	
Situation	Right	Wrong	Right	Wrong	Right	Wrong	Right	Wrong	Right	Wrong
1 (Algorithm 1)	20	2	20	–	20	1	20	1	20	2
1 (Tjoa)	19	–	20	1	20	–	19	1	20	2
2 (Algorithm 1)	10	–	20	–	20	–	19	–	20	–
2 (Tjoa)	12	–	20	–	20	1	20	1	17	1
3 (Algorithm 1)	20	1	20	2	20	2	20	2	20	1
3 (Tjoa)	20	2	20	–	20	–	19	–	20	1

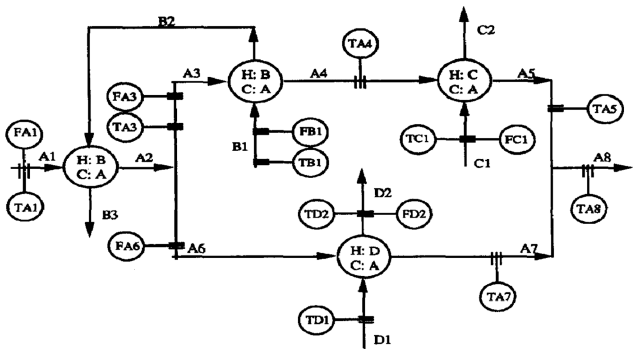

Figure 12. Flow sheet of the process for Swartz's example.[16]

of iterations of Algorithm 1 and the CONOPT solver. As an average, Algorithm 1 finishes in around 4 s within 10 iterations.

Conclusions

Gross error detection and data reconciliation problem is studied in this article. A unified framework is proposed to simultaneously estimate the true values of process variables x and u, the magnitudes of the gross errors δ, the covariance matrix Σ of the random error, and the gross error indicator matrix η. To reduce the solution complexity of directly solving the MAP problem, a hierarchical Bayesian framework is developed. The layered scheme allows us to obtain MAP estimates of the correct values of the measurements x, the magnitudes of the gross errors δ in the first layer, the covariance matrix Σ of the random error in the second layer and the gross error indicator matrix η in the third layer. To enhance the performance of the hierarchical Bayesian algorithm, the serial strategy is combined with the hierarchical Bayesian algorithm. The proposed method is not very sensitive to noise so that it can handle a large standard deviation of the noise. The proposed method is applicable to both linear and nonlinear cases. For linear case, the analytical forms of estimated values of the correct values of variables and magnitudes of gross errors are obtained. For nonlinear case, there is no need to do linearization or approximation in the proposed method.

The effectiveness of the proposed Bayesian approach and the serial strategies for simultaneous gross error detection and data reconciliation problem were demonstrated on simulated datasets. In the linear case, it is shown that the serial strategy is able to improve the performance under this situation. The results show that if the gross errors are correctly detected, the estimations of other variables will be very close to the exact values.

Table 7. Estimated Variables for Swartz's Example with TA7 as a Gross Error

Variable Tagname	Estimates of Variables (Tjoa)	Estimates of Variables (Algorithm 1)	Estimated σ (Algorithm 1)
FA1	968.38	963.84	21.75
TA1	466.33	466.31	0.8301
FA2	968.38	963.84	–
TA2	481.79	478.01	–
FA3	406.71	408.39	9.0927
TA3	481.79	478.01	0.6939
FA4	406.71	408.39	–
TA4	530.09	530.08	0.5892
FA5	406.71	408.39	–
TA5	616.21	613.58	0.6807
FA6	561.71	555.45	9.7999
TA6	481.79	478.01	–
FA7	561.67	555.45	–
TA7	614.38	617.79	0.6912
FA8	968.38	963.84	–
TA8	615.15	611.77	0.9952
FB1	253.20	253.30	4.8389
TB1	618.11	618.22	0.7844
FB2	253.20	253.30	–
TB2	543.86	545.76	–
FB3	253.20	253.30	–
TB3	486.58	492.65	–
FC1	308.10	307.02	8.0189
TC1	694.99	695.49	0.8827
FC2	308.10	307.02	–
TC2	594.21	605.98	–
FD1	680.88	688.01	
TD1	667.86	668.84	0.8884
FD2	680.88	688.01	12.4112
TD2	558.33	564.28	0.6693

Acknowledgment

The authors gratefully acknowledge the financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Notation

A = $p \times n$ incidence matrix
p = number of units in the network
n = number of flow rates
x = $n\times1$ vector of correct values of measured variables
D = $n \times m$ matrix of sample data
m = number of sampling points
y_{ij} = ith row, jth column entry of D
Y_j^C = jth column of D
Y_i^R = ith row of D
η = $n \times n$ diagonal indicator matrix
η_i = ith diagonal elements of η
δ = $n\times1$ vector of magnitudes of biases
δ_i = ith entry of δ
ε = Gaussian noise with mean zero and covariance Σ
Σ = $n \times n$ covariance diagonal matrix of noise
σ_{ii}^2 = ith diagonal element of Σ
α_i = inverse of σ_{ii}^2
μ_0 = $n\times1$ vector, mean hyperparameter of x
Σ_0 = $n \times n$ diagonal matrix, covariance hyperparameter of x
$\sigma_{ii,0}^2$ = ith diagonal element of Σ_0
α_{i0} = inverse of $\sigma_{ii,0}^2$
λ = $p\times1$ vector of Lagrange multiplier
s_i = inverse scale parameter of the prior distribution for α_i
k_i = shape parameter of the prior distribution for α_i
s_{i0} = inverse scale parameter of the prior distribution for α_{i0}
k_{i0} = shape parameter of the prior distribution for α_{i0}
p_i = the probability of the ith measurement containing a gross error
r_i, β_i = hyperparameters for the prior distribution of p_i
$\Gamma(.)$ = gamma function
I = indentity matrix

Literature Cited

1. Narasimhan S, Cornelius J. Data reconciliation and gross error detection: an intelligent use of process data. Houston, TX: Gulf Professional Publishing, 1999.
2. Mah RSH. *Chemical process structures and information flows*. Boston: Butterworths, 1990.
3. Reilly PM, Carpani RE. Application of statistical theory of adjustment to material balances. In: *Proceedings of the 13th Canadian Chemical Engineering Conference*. Montreal, Quebec, 1963.
4. Mah RSH, Tamhane AC. Detection of gross errors in process data. *AIChE J*. 1982;28(5):828–830.

5. Crowe CM, Campos YA, Hrymak A. Reconciliation of process flow rates by matrix projection. Part I: linear case. *AIChE J*. 1983;29(6): 881–888.
6. Mah RSH, Stanley GM, Downing DM. Reconcillation and rectification of process flow and inventory data. *Ind Eng Chem Process Des Dev*. 1976;15(1):175–183.
7. Narasimhan S, Mah RSH. Generalized likelihood ratio method for gross error identification. *AIChE J*. 1987;33(9):1514–1521.
8. Rollins DK, Davis JF. Unbiased estimation of gross errors in process measurements. *AIChE J*. 1992;38(4):563–572.
9. Tong H, Crowe CM. Detection of gross errors in data reconciliation by principal component analysis. *AIChE J*. 1995;41(7):1712–1722.
10. Ripps DL. Adjustment of experimental data. *Chem Eng Prog Symp Ser*. 1965;61:8–13.
11. Serth RW, Heenan WA. Gross error detection and data reconciliation in steam metering systems. *AIChE J*. 1986;32(5):733–742.
12. Rosenberg J, Mah RSH, Iordache C. Evaluation of schemes for detecting and identifying gross errors in process data. *Ind Eng Chem Res*. 1987;26(3):555–564.
13. Keller JY, Darouach M, Krzakala G. Fault detection of multiple biases or process leaks in linear steady state systems. *Comput Chem Eng*. 1994;18(10):1001–1004.
14. Sánchez M, Romagnoli J, Jiang Q, Bagajewicz M. Simultaneous estimation of biases and leaks in process plants. *Comput Chem Eng*. 1999;23(7):841–857.
15. Sánchez M, Romagnoli J. Use of orthogonal transformations in data classification-reconciliation. *Comput Chem Eng*. 1996;20(5):483–493.
16. Tjoa IB, Biegler LT. Simultaneous strategies for data reconciliation and gross error detection of nonlinear systems. *Comput Chem Eng*. 1991;15(10):679–690.
17. Soderstrom TA, Himmelblau DM, Edgar TF. A mixed integer optimization approach for simultaneous data reconciliation and identification of measurement bias. *Control Eng Pract*. 2001;9(8):869–876.
18. Han SP. A globally convergent method for nonlinear programming. *J Optim Theory Appl*. 1977;22(3):297–309.
19. Powell MJ. A fast algorithm for nonlinearly constrained optimization calculations. *Numerical Analysis*. G.A. Watson: Springer Berlin Heidelberg, 1978:144–157.
20. Chen HS, Stadtherr MA. Enhancements of the Han-Powell method for successive quadratic programming. *Comuput Chem Eng*. 1984; 8(3):229–234.
21. MacKay DJC. Bayesian interpolation. *Neural Comput*. 1992;4(3): 415–447.
22. MacKay DJC. Probable networks and plausible predictions-a review of practical Bayesian methods for supervised neural networks. *Network: Comput Neural Syst*. 1995;6(3):469–505.
23. Molina R, Vega M, Mateos J, Katsaggelos AK. Variational posterior distribution approximation in Bayesian super resolution reconstruction of multispectral images. *Appl Comput Harmon Anal*. 2008; 24(2):251–267.
24. Galatsanos NP, Mesarovic VZ, Molina R, Katsaggelos AK. Hierarchical Bayesian image restoration from partially known blurs. *IEEE Trans Image Process*. 2000;9(10):1784–1797.
25. Kwok JTY. The evidence framework applied to support vector machines. *IEEE Trans Neural Network*. 2000;11(5):1162–1173.
26. Suykens JAK, Gestel TV, Brabanter JD, Moor BD, Vandewalle J. *Least Squares Support Vector Machines*. River Edge, NJ: World Scientific, 2002.
27. Khatibisepehr S, Huang B. A Bayesian approach to robust process identification with ARX models. *AIChE J*. 2013;59(3):845–859.
28. Jiang Q, Bagajewicz MJ. On a strategy of serial identification with collective compensation for multiple gross error estimation in linear steady-state reconciliation. *Ind Eng Chem Res*. 1999;38(5):2119–2128.
29. Iordache C, Mah RSH, Tamhane AC. Performance studies of the measurement test for detection of gross errors in process data. *AIChE J*. 1985;31(7):1187–1201.
30. Bagajewicz MJ, Jiang Q. Gross error modeling and detection in plant linear dynamic reconciliation. *Comput Chem Eng*. 1998;22(12): 1789–1809.
31. Bagajewicz MJ, Jiang Q. A mixed integer linear programming-based technique for the estimation of multiple gross errors in process measurements. *Chem Eng Commun*. 2000;177(1):139–155.
32. Pai CCD, Fisher GD. Application of Broyden's method to reconciliation of nonlinearly constrained data. *AIChE J*. 1988;34:873–876.
33. Tamhane AC, Mah RSH. Data reconciliation and gross error detection in chemical process networks. *Technometrics*. 1985;27(4):409–422.
34. Crowe CM. Reconciliation of process flow rates by matrix projection. Part II: the nonlinear case. *AIChE J*. 1986;32(4):616–623.
35. Swartz CLE. Data reconciliation for generalized flowsheet applications. *Proceedings of the American Chemical Society National Meeting*. Dallas, TX, 1989.

Manuscript received Oct. 30, 2014, and revision received Feb. 20, 2015.